

PCTWORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification ⁶: C12N 15/60, 9/00, 9/88	A2	(11) International Publication Number: WO 95/10618 (43) International Publication Date: 20 April 1995 (20.04.95)
(21) International Application Number: PCT/EP94/03399 (22) International Filing Date: 15 October 1994 (15.10.94) (30) Priority Data: 9321301.5 15 October 1993 (15.10.93) GB (71) Applicant (for all designated States except US): DANISCO A/S [DK/DK]; Langebrogade 1, P.O. Box 17, DK-1001 Copenhagen K (DK). (72) Inventors; and (75) Inventors/Applicants (for US only): YU, Shukun [CN/SE]; Gunnar Hejdemans Gata 29, S-212 40 Malmo (SE). BO-JSEN, Kirsten [DK/DK]; Hvidehusvej 13, DK-3450 Allerod (DK). KRAGH, Karsten, Matthias [DK/DK]; Stavtrupvej 139 A, DK-8260 Viby J (DK). BOJKO, Maja [DK/DK]; Frugtparken 11, DK-2820 Gentofte (DK). NIELSEN, John [DK/DK]; Engvej 38, DK-2300 Copenhagen S (DK). MARCUSSEN, Jan [DK/DK]; Knabrostraede 25, 2., DK-1210 Copenhagen K (DK). (74) Agent: HARDING, Charles, Thomas; D. Young & Co., 21 New Fetter Lane, London EC4A 1DA (GB).		(81) Designated States: AM, AT, AU, BB, BG, BR, BY, CA, CH, CN, CZ, DE, DK, ES, FI, GB, GE, HU, JP, KE, KG, KP, KR, KZ, LK, LT, LU, LV, MD, MG, MN, MW, NL, NO, NZ, PL, PT, RO, RU, SD, SE, SI, SK, TJ, TT, UA, US, UZ, VN, European patent (AT, BE, CH, DE, DK, ES, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, ML, MR, NE, SN, TD, TG), ARIPO patent (KE, MW, SD, SZ). Published <i>Without international search report and to be republished upon receipt of that report.</i>
(54) Title: ALPHA-1,4-GLUCAN LYASE FROM A FUNGUS INFECTED ALGAE, ITS PURIFICATION, GENE CLONING AND EXPRESSION IN MICROORGANISMS (57) Abstract A method of preparing α -1,4-glucan lyase enzymes is described. The method comprises isolating the enzymes from a fungally infected algae. The amino acid sequences of the enzymes have been determined. The nucleic acid sequences coding for the enzymes have also been determined.		

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AT	Austria	GB	United Kingdom	MR	Mauritania
AU	Australia	GE	Georgia	MW	Malawi
BB	Barbados	GN	Guinea	NE	Niger
BE	Belgium	GR	Greece	NL	Netherlands
BF	Burkina Faso	HU	Hungary	NO	Norway
BG	Bulgaria	IE	Ireland	NZ	New Zealand
BJ	Benin	IT	Italy	PL	Poland
BR	Brazil	JP	Japan	PT	Portugal
BY	Belarus	KE	Kenya	RO	Romania
CA	Canada	KG	Kyrgyzstan	RU	Russian Federation
CF	Central African Republic	KP	Democratic People's Republic of Korea	SD	Sudan
CG	Congo	KR	Republic of Korea	SE	Sweden
CH	Switzerland	KZ	Kazakhstan	SI	Slovenia
CI	Côte d'Ivoire	LI	Liechtenstein	SK	Slovakia
CM	Cameroon	LK	Sri Lanka	SN	Senegal
CN	China	LU	Luxembourg	TD	Chad
CS	Czechoslovakia	LV	Latvia	TG	Togo
CZ	Czech Republic	MC	Monaco	TJ	Tajikistan
DE	Germany	MD	Republic of Moldova	TT	Trinidad and Tobago
DK	Denmark	MG	Madagascar	UA	Ukraine
ES	Spain	ML	Mali	US	United States of America
FI	Finland	MN	Mongolia	UZ	Uzbekistan
FR	France			VN	Viet Nam
GA	Gabon				

ALPHA-1,4-GLUCAN LYASE FROM A FUNGUS INFECTED ALGAE, ITS PURIFICATION, GENE CLONING AND EXPRESSION IN MICROORGANISMS

The present invention relates to an enzyme, in particular α -1,4-glucan lyase ("GL"). The present invention also relates to a method of extracting the same. The present invention also relates to nucleotide sequence(s) encoding for the same.

FR-A-2617502 and Baute et al in Phytochemistry [1988] vol. 27 No.11 pp3401-3403 report on the production of 1,5-D-anhydrofructose ("AF") in *Morchella vulgaris* by an apparent enzymatic reaction. The yield of production of AF is quite low. Despite a reference to a possible enzymatic reaction, neither of these two documents presents any amino acid sequence data for any enzyme, let alone any nucleotide sequence information. These documents say that AF can be a precursor for the preparation of the antibiotic pyrone microthecin.

Yu et al in Biochimica et Biophysica Acta [1993] vol 1156 pp313-320 report on the preparation of GL from red seaweed and its use to degrade α -1,4-glucan to produce AF. The yield of production of AF is quite low. Despite a reference to the enzyme GL this document does not present any amino acid sequence data for that enzyme let alone any nucleotide sequence information coding for the same. This document also suggests that the source of GL is just algal.

According to the present invention there is provided a method of preparing the enzyme α -1,4-glucan lyase comprising isolating the enzyme from a fungally infected algae.

Preferably the enzyme is isolated and/or further purified using a gel that is not degraded by the enzyme.

Preferably the gel is based on dextrin, preferably beta-cyclodextrin, or derivatives thereof, preferably a cyclodextrin, more preferably beta-cyclo-dextrin.

According to the present invention there is also provided a GL enzyme prepared by the method of the present invention.

5 Preferably the enzyme comprises the amino acid sequence SEQ. ID. No. 1. or SEQ. ID. No. 2, or any variant thereof.

The term "any variant thereof" means any substitution of, variation of, modification of, replacement of, deletion of or addition of at least one amino acid from or to the sequence providing the resultant enzyme has lyase activity.

10

According to the present invention there is also provided a nucleotide sequence coding for the enzyme α -1,4-glucan lyase, preferably wherein the sequence is not in its natural environment (i.e. does not form part of the natural genome of a cellular organism expressing the enzyme).

15

Preferably the nucleotide sequence is a DNA sequence.

Preferably the DNA sequence comprises a sequence that is the same as, or is complementary to, or has substantial homology with, or contains any suitable codon substitution(s) for any of those of, SEQ. ID. No. 3 or SEQ. ID. No. 4.

20

The expression "substantial homology" covers homology with respect to structure and/or nucleotide components and/or biological activity.

25 The expression "contains any suitable codon substitutions" covers any codon replacement or substitution with another codon coding for the same amino acid or any addition or removal thereof providing the resultant enzyme has lyase activity.

30 In other words, the present invention also covers a modified DNA sequence in which at least one nucleotide has been deleted, substituted or modified or in which at least one additional nucleotide has been inserted so as to encode a polypeptide having the activity of a glucan lyase, preferably an enzyme having an increased lyase activity.

According to the present invention there is also provided a method of preparing the enzyme α -1,4-glucan lyase comprising expressing the nucleotide sequence of the present invention.

5 According to the present invention there is also provided the use of beta-cyclodextrin to purify an enzyme, preferably GL.

10 According to the present invention there is also provided a nucleotide sequence wherein the DNA sequence comprises a sequence that is the same as, or is complementary to, or has substantial homology with, or contains any suitable codon substitutions for any of those of, SEQ. ID. No. 3 or SEQ. ID. No. 4, preferably wherein the sequence is in isolated form.

15 A key aspect of the present invention is the recognition that GL is derived from a fungally infected algae. This is the first time that the amino acid sequence of GL has been determined in addition to the determination of the nucleic acid sequences that code for GL. A key advantage of the present invention is therefore that GL can now be made in large quantities by for example recombinant DNA techniques and thus enable compounds such as the antibiotic microthecin to be made easily and in larger amounts.

20

The enzyme should preferably be secreted to ease its purification. To do so the DNA encoding the mature enzyme is fused to a signal sequence, a promoter and a terminator from the chosen host.

25

For expression in *Aspergillus niger* the *gpdA* (from the Glyceraldehyde-3-phosphate dehydrogenase gene of *Aspergillus nidulans*) promoter and signal sequence is fused to the 5' end of the DNA encoding the mature lyase - such as SEQ I.D. No. 3 or SEQ. I.D. No.4. The terminator sequence from the *A. niger* *trpC* gene is placed 3' to the gene (Punt, P.J. et al (1991): J. Biotech. 17, 19-34). This construction is inserted into a vector containing a replication origin and selection origin for *E. coli* and a selection marker for *A. niger*. Examples of selection markers for *A. niger* are

30

the amdS gene, the argB gene, the pyrG gene, the hygB gene, the BmlR gene which all have been used for selection of transformants. This plasmid can be transformed into *A. niger* and the mature lyase can be recovered from the culture medium of the transformants.

5

The construction can be transformed into a protease deficient strain to reduce the proteolytic degradation of the lyase in the culture medium (Archer D.B. et al (1992): Biotechnol. Lett. 14, 357-362).

10

Other advantages will become apparent in the light of the following description.

15

The present invention therefore relates to the isolation of the enzyme α -1,4-glucan lyase from a fungus infected algae - preferably a fungus infected red algae such as the type that can be collected in China - such as *Gracilariopsis lemaneiformis*. An example of a fungally infected algae has been deposited in accordance with the Budapest Treaty (see below).

20

By using *in situ* hybridisation technique it was established that the enzyme GL was detected in the fungally infected red algae *Gracilariopsis lemaneiformis*. Further evidence that supports this observation was provided by the results of Southern hybridisation experiments. Thus GL enzyme activity can be obtained from fungally infected algae, rather than just from the algae as was originally thought.

25

Of particular interest is the finding that there are two natural DNA sequences, each of which codes for an enzyme having GL characteristics. These DNA nucleic acid sequences have been sequenced and they are presented as SEQ. I.D. No. 3 and SEQ. I.D. No. 4 (which are discussed and presented later).

30

An initial enzyme purification can be performed by the method as described by Yu et al (ibid). However, it is preferred that the initial enzyme purification includes the use of a solid support that does not decompose under the purification step. This gel support has the advantage that it is compatible with standard laboratory protein

purification equipment. The details of this preferred purification process are given later on. The purification is terminated by known standard techniques for protein purification. The purity of the enzyme was established using complementary electroforetic techniques.

5

The purified lyase was characterized according to pI, temperature- and pH-optima. In this regard, it was found that the enzyme has the following characteristics: an optimum substrate specificity and a pH optimum at 3.5-7.5 when amylopectin is used; a temperature optimum at 50°C and a pI of 3.9

10

As mentioned above, the enzymes according to the present invention have been determined (partially by amino-acid sequencing techniques) and their amino acid sequences are provided later. Likewise the nucleotide sequences coding for the enzymes according to the present invention (i.e. GL) have been sequenced and the DNA sequences are provided later.

15

The following samples were deposited in accordance with the Budapest Treaty at the recognised depositary The National Collections of Industrial and Marine Bacteria Limited (NCIMB) at 23 St. Machar Drive, Aberdeen, Scotland, United Kingdom, AB2 1RY on 20 June 1994:

20

E. Coli containing plasmid pGL1 (NCIMB 40652) - [ref. DH5alpha-pGL1]; and

E. Coli containing plasmid pGL2 (NCIMB 40653) - [ref. DH5alpha-pGL2].

25

The following sample was accepted as a deposit in accordance with the Budapest Treaty at the recognised depositary The Culture Collection of Algae and Protozoa (CCAP) at Dunstaffnage Marine Laboratory PO Box 3, Oban, Argyll, Scotland, United Kingdom, PA34 4AD on 11 October 1994:

30

Fungally infected *Gracilariopsis lemaneiformis* (CCAP 1373/1) - [ref. GLQ-1 (Qingdao)].

Thus highly preferred embodiments of the present invention include a GL enzyme obtainable from the expression of the GL coding sequences present in plasmids that are the subject of either deposit NCIMB 40652 or deposit NCIMB 40653; and a GL enzyme obtainable from the fungally infected algae that is the subject of deposit
5 CCAP 1373/1.

The present invention will now be described only by way of example.

In the following Examples reference is made to the accompanying figures in which:
10

Figure 1 shows stained fungally infected algae;

Figure 2 shows stained fungally infected algae;

15 Figure 3 shows sections of fungal hypha;

Figure 4 shows sections of fungally infected algae;

Figure 5 shows a section of fungally infected algae;
20

Figure 6 shows a plasmid map of pGL1;

Figure 7 shows a plasmid map of pGL2;

25 Figure 8 shows the amino acid sequence represented as SEQ. I.D. No.3 showing positions of the peptide fragments that were sequenced;

Figure 9 shows the alignment of SEQ. I.D. No. 1 with SEQ. I.D. No.2;

30 Figure 10 is a microphotograph.

In more detail, Figure 1 shows Calcoflour White stainings revealing fungi in upper part and lower part of *Gracilariopsis lemaneiformis* (108x and 294x).

5 Figure 2 shows PAS/Anilinblue Black staining of *Gracilariopsis lemaneiformis* with fungi. The fungi have a significant higher content of carbohydrates.

10 Figure 3 shows a micrograph showing longitudinal and grazing sections of two thin-walled fungal hypha (f) growing between thick walls (w) of algal cells. Note thylacoid membranes in the algal chloroplast (arrows).

Figure 4 shows the antisense detections with clone 2 probe (upper row) appear to be restricted to the fungi illustrated by Calcoflour White staining of the succeeding section (lower row) (46x and 108x).

15 Figure 5 shows intense antisense detections with clone 2 probe are found over the fungi in *Gracilariopsis lemaneiformis* (294x).

20 Figure 6 shows a map of plasmid pGL1 - which is a pBluescript II KS containing a 3.8 kb fragment isolated from a genomic library constructed from fungal infected *Gracilariopsis lemaneiformis*. The fragment contains a gene coding for alpha-1,4-glucan lyase.

25 Figure 7 shows a map of plasmid pGL2 - which is a pBluescript II SK containing a 3.6 kb fragment isolated from a genomic library constructed from fungal infected *Gracilariopsis lemaneiformis*. The fragment contains a gene coding for alpha-1,4-glucan lyase.

30 Figure 9 shows the alignment of SEQ. I.D. No. 1 (GL1) with SEQ. I.D. No.2 (GL2). The total number of residues for GL1 is 1088; and the total number of residues for GL2 is 1091. In making the comparison, a structure-genetic matrix was used (Open gap cost: 10; Unit gap cost: 2). In Figure 9 the character to show that two aligned residues are identical is ':'; and the character to show that two aligned

residues are similar is '. '. Amino acids said to be 'similar' are: A,S,T; D,E; N,Q; R,K; I,L,M,V; F,Y,W. Overall there is an identity of 845 amino acids (i.e. 77.67%); a similarity of 60 amino acids (5.51%). The number of gaps inserted in GL1 are 3 and the number of gaps inserted in GL2 are 2.

5

Figure 10 is a microphotograph of a fungal hypha (f) growing between the algal walls (w). Note grains of floridean starch (s) and thylakoids (arrows) in the algal cell.

The following sequence information was used to generate primers for the PCR reactions mentioned below and to check the amino acid sequence generated by the respective nucleotide sequences.

10

Amino acid sequence assembled from peptides from fungus infected *Gracilariopsis lemaneiformis*

15

Tyr Arg Trp Gln Glu Val Leu Tyr Thr Ala Met Tyr Gln Asn Ala
Ala Phe Gly Lys Pro Ile Ile Lys Ala Ala Ser Met Tyr Asn Asn
Asp Ser Asn Val Arg Arg Ala Gln Asn Asp His Phe Leu Leu Gly
Gly His Asp Gly Tyr Arg Ile Leu Cys Ala Pro Val Val Trp Glu
Asn Ser Thr Glu Arg Glu Leu Tyr Leu Pro Val Leu Thr Gln Trp
Tyr Lys Phe Gly Pro Asp Phe Asp Thr Lys Pro Leu Glu Gly Ala

20

The Amino acid sequence (27-34) used to generate primer A and B (Met Tyr Asn Asn Asp Ser Asn Val)

25

Primer A

ATG TA(TC) AA(CT) AA(CT) GA(CT) TC(GATC) AA(CT) GT 128 mix

Primer B

30

ATG TA(TC) AA(CT) AA(CT) GA(CT) AG(CT) AA(CT) GT 64 mix

The Amino acid sequence (45-50) used to generate primer C (Gly Gly His Asp Gly Tyr)

Primer C

5 TA (GATC)CC (GA)TC (GA)TG (GATC)CC (GATC)CC 256 mix
[The sequence corresponds to the complementary strand.]

The Amino acid sequence (74-79) used to generate primer E (Gln Trp Tyr Lys Phe Gly)

10

Primer E

GG(GATC) CC(GA) AA(CT) TT(GA) TAC CA(CT) TG 64 mix
[The sequence corresponds to the complementary strand.]

15 The Amino acid sequence (1-6) used to generate primer F1 and F2 (Tyr Arg Trp Gln Glu Val)

Primer F1

20 TA(TC) CG(GATC) TGG CA(GA) GA(GA) GT 32 mix

Primer F2

TA(TC) AG(GA) TGG CA(GA) GA(GA) GT 16 mix

The sequence obtained from the first PCR amplification (clone 1)

25

ATGTACAACA ACGACTCGAA CGTTCGCAGG GCGCAGAACG ATCATTTCCT
TCTTGCGGC CACGACGGTT A

30 Met Tyr Asn Asn Asp Ser Asn Val Arg Arg Ala Gln Asn Asp His Phe Leu Leu Gly
Gly His Asp Gly

The sequence obtained from the second PCR amplification (clone 1)

ATGTACAACA ACGACTCGAA CGTTCGCAGG GCGCAGAACG ATCATTTCCT
TCTTGGTGGA CATGATGGAT ATCGCATTCT GTGCGCGCCT GTTGTGTGGG
AGAATTCGAC CGAACGNGAA TTGTACTTGC CCGTGCTGAC CCAATGGTAC
5 AAATTCGGCC C

Met Tyr Asn Asn Asp Ser Asn Val Arg Arg Ala Gln Asn Asp His Phe Leu Leu Gly
Gly His Asp Gly Tyr Arg Ile Leu Cys Ala Pro Val Val Trp Glu Asn Ser Thr Glu
Arg Glu Leu Tyr Leu Pro Val Leu Thr Gln Trp Tyr Lys Phe Gly Pro

10

The sequence obtained from the third PCR amplification (clone2)

TACAGGTGGC AGGAGGTGTT GTACACTGCT ATGTACCAGA
ATGCGGCTTT CGGGAAACCG ATTATCAAGG CAGCTTCCAT
15 GTACGACAAC GACAGAAACG TTCGCGGCGC ACAGGATGAC
CACTTCCTTC TCGGCGGACA CGATGGATAT CGTATTTTGT
GTGCACCTGT TGTGTGGGAG AATACAACCA GTCGCGATCT
GTACTTGCCT GTGCTGACCA GTGGTACAAA TTCGGCCC

20 Tyr Arg Trp Gln Glu Val Leu Tyr Thr Ala Met Tyr Gln Asn Ala Ala Phe Gly Lys
Pro Ile Ile Lys Ala Ala Ser Met Tyr Asp Asn Asp Arg Asn Val Arg Gly Ala Gln Asp
Asp His Phe Leu Leu Gly Gly His Asp Gly Tyr Arg Ile Leu Cys Ala Pro Val Val
Trp Glu Asn Thr Thr Ser Arg Asp Leu Tyr Leu Pro Val Leu Thr Lys Trp Tyr Lys
Phe Gly

25

1. CYTOLOGICAL INVESTIGATIONS OF GRACILARIOPSIS LEMANEIFORMIS

1.1.1 Detection of fungal infection in *Gracilariopsis lemaneiformis*

Sections of *Gracilariopsis lemaneiformis* collected in China were either hand cut or cut from paraffin embedded material. Sectioned material was carefully investigated by light microscopy. Fungal hyphae were clearly detected in *Gracilariopsis lemaneiformis*.

The thalli of the *Gracilariopsis lemaneiformis* are composed of cells appearing in a highly ordered and almost symmetric manner. The tubular thallus of *G. lemaneiformis* is composed of large, colourless central cells surrounded by elongated, slender, elliptical cells and small, round, red pigmented peripheral cells. All algal cell types are characterized by thick cell walls. Most of the fungal hyphae are found at the interphase between the central layer of large cells and the peripheral layer. These cells can clearly be distinguished from the algae cells as they are long and cylindrical. The growth of the hyphae is observed as irregularities between the highly ordered algae cells. The most frequent orientation of the hypha is along the main axis of the algal thallus. Side branches toward the central and periphery are detected in some cases. The hypha can not be confused with the endo/epiphytic 2nd generation of the algae.

Calcofluor White is known to stain chitin and cellulose containing tissue. The reaction with chitin requires four covalently linked terminal n-acetyl glucosamine residues. It is generally accepted that cellulose is almost restricted to higher plants although it might occur in trace amounts in some algae. It is further known that chitin is absent in *Gracilaria*.

Calcofluor White was found to stain domains corresponding to fungi hypha cell walls in sectioned *Gracilariopsis lemaneiformis* material.

The hypha appear clear white against a faint blue background of *Gracilaria* tissue when observed under u.v. light - see Figure 1. Chitin is the major cell wall component in most fungi but absent in *Gracilaria*. Based upon these observations we conclude that the investigated algae is infected by a fungi. 40% of the lower parts of the investigated *Gracilariopsis lemaneiformis* sections were found to be infected with fungal hyphae. In the algae tips 25% of the investigated *Gracilariopsis lemaneiformis* sections were found to be infected.

Staining of sectioned *Gracilariopsis lemaneiformis* with Periodic acid Schiff (PAS) and Aniline blue black revealed a significantly higher content of carbohydrates within the fungal cells as compared with the algae cells - see Figure 2. Safranin O and Malachit Green showed the same colour reaction of fungi cells as found in higher plants infected with fungi.

An Acridin Orange reaction with sectioned *Gracilariopsis lemaneiformis* showed clearly the irregularly growth of the fungus.

1.1.2 Electron Microscopy

Slides with 15 μm thick sections, where the fungus was detected with Calcofluor White were fixed in 2% OsO_4 , washed in water and dehydrated in dimethoxypropane and absolute alcohol. A drop of a 1:1 mixture of acetone and Spurr resin was placed over each section on the glass slide, and after one hour replaced by a drop of pure resin. A gelatin embedding capsule filled with resin was placed face down over the section and left over night at 4°C. After the polymerization at 55°C for 8 hrs, the thick sections adhering to the resin blocks could be separated from the slide by immersion in liquid nitrogen.

Blocks were trimmed and 100 nm thick sections were cut using a diamond knife on a microtome. The sections were stained in aqueous uranyl acetate and in lead citrate. The sections were examined in an electron microscope at 80 kV.

The investigation confirmed the light microscopical observations and provided further evidence that the lyase producing, chinese strain of *G. lamneiformis* is infected by a fungal parasite or symbiont.

5 Fungal hyphae are build of tubular cells 50 to 100 μm long and only few microns in diameter. The cells are serially arranged with septate walls between the adjacent cells. Ocasional branches are also seen. The hyphae grow between the thick cell walls of algal thallus without penetrating the wall or damaging the cell. Such a symbiotic association, called mycophycobiosis, is known to occur between some filamentous
10 marine fungi and large marine algae (Donk and Bruning, 1992 - Ecology of aquatic fungi in and on algae. In Reisser, W.(ed.): Algae and Symbioses: Plants, Animals, Fungi, Viruses, Interactions Explored. Biopress Ltd.,Bristol.)

15 Examining the microphotograph in Figure 10, several differences between algal and fungal cells can be noticed. In contrast to several μm thick walls of the alga, the fungal walls are only 100-200 nm thick. Plant typical organells as chloroplasts with thyllacoid membranes as well as floridean starch grains can be seen in algal cells, but not in the fungus.

20 Intercellular connections of red algae are characterized by specific structures termed pit plugs, or pit connections. The structures are prominent, electron dense cores and they are important features in algal taxonomy (Pueschel, C.M.: An expanded survey of the ultrastructure of Red algal pit plugs. J. Phycol. 25, 625, (1989)). In our material, such connections were frequently observed in the algal thallus, but never
25 between the cells of the fungus.

1.2 *In situ* Hybridization experiments

30 *In situ* hybridization technique is based upon the principle of hybridization of an antisense ribonucotide sequence to the mRNA. The technique is used to visualize areas in microscopic sections where said mRNA is present. In this particular case the technique is used to localize the enzyme α -1,4-glucan lyase in sections of

Gracilariopsis lemaneiformis.1.2.1 Preparation of ^{35}S labelled probes for *In situ* hybridization

5 A 238 bp PCR fragment from a third PCR amplification - called clone 2 (see above) - was cloned into the pGEM-3Zf(+) Vector (Promega). The transcription of the antisense RNA was driven by the SP6 promotor, and the sense RNA by the T7 promotor. The Ribonuclease protection assay kit (Ambion) was used with the following modifications. The transcripts were run on a 6% sequencing gel to remove
10 the unincorporated nucleotide and eluted with the elution buffer supplied with the T7RNA polymerase in vitro Transcription Kit (Ambion). The antisense transcript contained 23 non-coding nucleotides while the sense contained 39. For hybridization 10^7 cpm/ml of the ^{35}S labelled probe was used.

15 *In situ* hybridisation was performed essentially as described by Langedale et.al.(1988). The hybridization temperature was found to be optimal at 45°C. After washing at 45°C the sections were covered with Kodak K-5 photographic emulsion and left for 3 days at 5°C in dark (Ref: Langedale, J.A., Rothermel, B.A. and Nelson, T. (1988). Genes and development 2: 106-115. Cold Spring Harbour Laboratory).
20

The *in situ* hybridization experiments with riboprobes against the mRNA of α -1,4-glucan lyase, show strong hybridizations over and around the hypha of the fungus detected in *Gracilariopsis lemaneiformis* - see Figures 4 and 5. This is considered
25 a strong indication that the α -1,4-glucan lyase is produced. A weak random background reactions were detected in the algae tissue of both *Gracilariopsis lemaneiformis*. This reaction was observed both with the sense and the antisense probes. Intense staining over the fungi hypha was only obtained with antisense probes.

30

These results were obtained with standard hybridisation conditions at 45°C in hybridization and washing steps. At 50°C no staining over the fungi was observed,

whereas the background staining remained the same. Raising the temperature to 55°C reduced the background staining with both sense and antisense probes significantly and equally.

5 Based upon the cytological investigations using complementary staining procedures it is concluded that *Gracilariopsis lemaneiformis* is fungus infected. The infections are most pronounced in the lower parts of the algal tissue.

10 In sectioned *Gracilariopsis lemaneiformis* material *in situ* hybridization results clearly indicate that hybridization is restricted to areas where fungal infections are found - see Figure 4. The results indicate that α -1,4-glucan lyase mRNA appears to be restricted to fungus infected areas in *Gracilariopsis lemaneiformis*.

15 Based upon these observations we conclude that α -1,4-glucan lyase activity is detected in fungally infected *Gracilariopsis lemaneiformis*.

2. ENZYME PURIFICATION AND CHARACTERIZATION

20 Purification of α -1,4-glucan lyase from fungal infected *Gracilariopsis lemaneiformis* material was performed as follows.

2.1 Materials and Methods

25 The algae were harvested by filtration and washed with 0.9% NaCl. The cells were broken by homogenization followed by sonication on ice for 6x3 min in 50 mM citrate-NaOH pH 6.2 (Buffer A). Cell debris were removed by centrifugation at 25,000xg for 40 min. The supernatant obtained at this procedure was regarded as cell-free extract and was used for activity staining and Western blotting after separation on 8-25% gradient gels.

30

2.2 Separation by β -cyclodextrin Sepharose gel

5 The cell-free extract was applied directly to a β -cyclodextrin Sepharose gel 4B column (2.6 x 18 cm) pre equilibrated with Buffer A. The column was washed with 3 volumes of Buffer A and 2 volumes of Buffer A containing 1 M NaCl. α -1,4-glucan lyase was eluted with 2 % dextrans in Buffer A. Active fractions were pooled and the buffer changed to 20 mM Bis-tris propane-HCl (pH 7.0, Buffer B).

10 Active fractions were applied onto a Mono Q HR 5/5 column pre-equilibrated with Buffer B. The fungal lyase was eluted with Buffer B in a linear gradient of 0.3 M NaCl.

15 The lyase preparation obtained after β -cyclodextrin Sepharose chromatography was alternatively concentrated to 150 μ l and applied on a Superose 12 column operated under FPLC conditions.

2.3 Assay for α -1,4-glucan lyase activity and conditions for determination of substrate specificity, pH and temperature optimum

20 The reaction mixture for the assay of the α -1,4-glucan lyase activity contained 10 mg ml⁻¹ amylopectin and 25 mM Mes-NaOH (pH 6.0). The reaction was carried out at 30°C for 30 min and stopped by the addition of 3,5-dinitrosalicylic acid reagent. Optical density at 550nm was measured after standing at room temperature for 10 min.

25

3. AMINO ACID SEQUENCING OF THE α -1,4-GLUCAN LYASE FROM FUNGUS INFECTED GRACILARIOPSIS LEMANEIFORMIS

3.1 Amino acid sequencing of the lyases

5

The lyases were digested with either endoproteinase Arg-C from *Clostridium histolyticum* or endoproteinase Lys-C from *Lysobacter enzymogenes*, both sequencing grade purchased from Boehringer Mannheim, Germany. For digestion with endoproteinase Arg-C, freeze dried lyase (0.1 mg) was dissolved in 50 μ l 10 M urea, 50 mM methylamine, 0.1 M Tris-HCl, pH 7.6. After overlay with N₂ and addition of 10 μ l of 50 mM DTT and 5 mM EDTA the protein was denatured and reduced for 10 min at 50°C under N₂. Subsequently, 1 μ g of endoproteinase Arg-C in 10 μ l of 50 mM Tris-HCl, pH 8.0 was added, N₂ was overlayed and the digestion was carried out for 6h at 37°C. For subsequent cysteine derivatization, 12.5 μ l 100 mM iodoacetamide was added and the solution was incubated for 15 min at RT in the dark under N₂.

20

For digestion with endoproteinase Lys-C, freeze dried lyase (0.1 mg) was dissolved in 50 μ l of 8 M urea, 0.4 M NH₄HCO₃, pH 8.4. After overlay with N₂ and addition of 5 μ l of 45 mM DTT, the protein was denatured and reduced for 15 min at 50°C under N₂. After cooling to RT, 5 μ l of 100 mM iodoacetamide was added for the cysteines to be derivatized for 15 min at RT in the dark under N₂.

25

Subsequently, 90 μ l of water and 5 μ g of endoproteinase Lys-C in 50 μ l of 50 mM tricine and 10 mM EDTA, pH 8.0, was added and the digestion was carried out for 24h at 37°C under N₂.

30

The resulting peptides were separated by reversed phase HPLC on a VYDAC C18 column (0.46 x 15 cm; 10 μ m; The Separations Group; California) using solvent A: 0.1% TFA in water and solvent B: 0.1% TFA in acetonitrile. Selected peptides were rechromatographed on a Develosil C18 column (0.46 x 10 cm; 3 μ m; Dr. Ole Schou, Novo Nordisk, Denmark) using the same solvent system prior to sequencing on an

Applied Biosystems 476A sequencer using pulsed-liquid fast cycles.

The amino acid sequence information from the enzyme derived from fungus infected *Gracilariopsis lemaneiformis* is shown below, in particular SEQ. ID. No. 1. and
5 SEQ. ID. No. 2.

SEQ. I.D. No. 1 has:

Number of residues : 1088.

Amino acid composition (including the signal sequence)

10

=====

61 Ala	15 Cys	19 His	34 Met	78 Thr
51 Arg	42 Gln	43 Ile	53 Phe	24 Trp
88 Asn	53 Glu	63 Leu	51 Pro	58 Tyr
79 Asp	100 Gly	37 Lys	62 Ser	77 Val

15

SEQ. I.D. No. 2 has:

Number of residues : 1091.

Amino acid composition (including the signal sequence)

=====

20

58 Ala	16 Cys	14 His	34 Met	68 Thr
57 Arg	40 Gln	44 Ile	56 Phe	23 Trp
84 Asn	47 Glu	69 Leu	51 Pro	61 Tyr
81 Asp	102 Gly	50 Lys	60 Ser	76 Val

25

3.2 N-TERMINAL ANALYSIS

Studies showed that the N-terminal sequence of native glucan lyase 1 was blocked. Deblocking was achieved by treating glucan lyase 1 blotted onto a PVDF membrane with anhydrous TFA for 30 min at 40°C essentially as described by LeGendre et al.
30 (1993) [Purification of proteins and peptides by SDS-PAGE; In: Matsudaira, P. (ed.) A practical guide to protein and peptide purification for microsequencing, 2nd edition; Academic Press Inc., San Diego; pp. 74-101.]. The sequence obtained was

TALSDKQTA, which matches the sequence (sequence position from 51 to 59 of SEQ. I.D. No.1) derived from the clone for glucan lyase 1 and indicates N-acetylthreonine as N-terminal residue of glucan lyase 1. Sequence position 1 to 50 of SEQ. I.D. No. 1 represents a signal sequence.

5

4. DNA SEQUENCING OF GENES CODING FOR THE α -1,4-GLUCAN LYASE FROM FUNGUS INFECTED GRACILARIOPSIS LEMANEIFORMIS

4.1 METHODS FOR MOLECULAR BIOLOGY

10

DNA was isolated as described by Saunders (1993) with the following modification: The polysaccharides were removed from the DNA by ELUTIP-d (Schleicher & Schuell) purification instead of gel purification. (Ref:Saunders, G.W. (1993). Gel purification of red algal genomic DNA: An inexpensive and rapid method for the isolation of PCR-friendly DNA. Journal of phycology 29(2): 251-254 and Schleicher & Schuell: ELUTIP-d. Rapid Method for Purification and Concentration of DNA.)

15

4.2 PCR

The preparation of the relevant DNA molecule was done by use of the Gene Amp DNA Amplification Kit (Perkin Elmer Cetus, USA) and in accordance with the manufactures instructions except that the Taq polymerase was added later (see PCR cycles) and the temperature cycling was changed to the following:

20

PCR cycles:

25

no of cycles	C	time (min.)
1	98	5
	60	5
addition of Taq polymerase and oil		
35	94	1
	47	2
	72	3
1	72	20

30

4.3 CLONING OF PCR FRAGMENTS

PCR fragments were cloned into pT7Blue (from Novagen) following the instructions of the supplier.

5

4.4 DNA SEQUENCING

Double stranded DNA was sequenced essentially according to the dideoxy method of Sanger et al. (1979) using the Auto Read Sequencing Kit (Pharmacia) and the Pharmacia LKB A.L.F.DNA sequencer. (Ref: Sanger, F., Nicklen, S. and Coulson, A.R.(1979). DNA sequencing with chain-determinating inhibitors. Proc. Natl. Acad. Sci. USA 74: 5463-5467.).

10

The sequences are shown as SEQ. I.D.No.s 3 and 4, wherein

15

SEQ. I.D. No. 3 has:

Total number of bases is: 3267.

DNA sequence composition: 850 A; 761 C; 871 G; 785 T

20

SEQ. I.D. No. 4 has:

Total number of bases is: 3276.

DNA sequence composition: 889 A; 702 C; 856 G; 829 T

4.5 SCREENING OF THE LIBRARY

25

Screening of the Lambda Zap library obtained from Stratagene, was performed in accordance with the manufacturer's instructions except that the prehybridization and hybridization was performed in 2xSSC, 0.1% SDS, 10xDenhardt's and 100µg/ml denatured salmon sperm DNA. To the hybridization solution a 32P-labeled denatured probe was added. Hybridization was performed over night at 55°C. The filters were washed twice in 2xSSC, 0.1% SDS and twice in 1xSSC, 0.1% SDS.

30

4.6 PROBE

The cloned PCR fragments were isolated from the pT7 blue vector by digestion with appropriate restriction enzymes. The fragments were separated from the vector by agarose gel electrophoresis and the fragments were purified from the agarose by Agarase (Boehringer Mannheim). As the fragments were only 90-240 bp long the isolated fragments were exposed to a ligation reaction before labelling with ^{32}P -dCTP using either Prime-It random primer kit (Stratagene) or Ready to Go DNA labelling kit (Pharmacia).

4.7 RESULTS

4.7.1 Generation of PCR DNA fragments coding for α -1,4-glucan lyase.

The amino acid sequences of three overlapping tryptic peptides from α -1,4-glucan lyase were used to generate mixed oligonucleotides, which could be used as PCR primers (see the sequences given above).

In the first PCR amplification primers A/B (see above) were used as upstream primers and primer C (see above) was used as downstream primer. The size of the expected PCR product was 71 base pairs.

In the second PCR amplification primers A/B were used as upstream primers and E was used as downstream primer. The size of the expected PCR product was 161 base pairs.

In the third PCR amplification primers F1 (see above) and F2 (see above) were used as upstream primers and E was used as downstream primer. The size of the expected PCR product was 238 base pairs. The PCR products were analysed on a 2% LMT agarose gel and fragments of the expected sizes were cut out from the gel and treated with Agarase (Boehringer Mannheim) and cloned into the pT7blue Vector (Novagen) and sequenced.

The cloned fragments from the first and second PCR amplification coded for amino acids corresponding to the sequenced peptides (see above). The clone from the third amplification (see above) was only about 87% homologous to the sequenced peptides.

5 4.7.2 Screening of the genomic library with the cloned PCR fragments.

Screening of the library with the above-mentioned clones gave two clones. One clone contained the nucleotide sequence of SEQ I.D. No. 4 (gene 2). The other clone contained some of the sequence of SEQ I.D. No.3 (from base pair 1065 downwards)
10 (gene 1).

The 5' end of SEQ. I.D. No. 3 (i.e. from base pair 1064 upwards) was obtained by the RACE (rapid amplification of cDNA ends) procedure (Michael, A.F., Michael, K.D. & Martin, G.R.(1988). Proc..Natl.Acad.Sci.USA 85:8998-99002.) using the
15 5' race system from Gibco BRL. Total RNA was isolated according to Collinge et al.(Collinge, D.B., Milligan D.E., Dow, J.M., Scofield, G.& Daniels, M.J.(1987). Plant Mol Biol 8: 405-414). The 5' race was done according to the protocol of the manufacturer, using 1µg of total RNA. The PCR product from the second
20 ammplification was cloned into pT7blue vector from Novagen according to the protocol of the manufacturer. Three independent PCR clones were sequenced to compensate for PCR errors.

An additional PCR was performed to supplement the clone just described with XbaI and NdeI restriction sites immediately in front of the ATG start codon using the
25 following oligonucleotide as an upstream primer:

GCTCTAGAGCATGTTTTC AACCCTTGCG and a primer containing the complement sequence of bp 1573-1593 in sequence GLI (i.e. SEQ. I.D. No. 3) was used as a downstream primer.

30 The complete sequence for gene 1 (i.e. SEQ. I.D. No. 3) was generated by cloning the 3' end of the gene as a BamHI-HindIII fragment from the genomic clone into the pBluescript II KS+ vector from Stratagene and additionally cloning the PCR

generated 5' end of the gene as a XbaI-BamHI fragment in front of the 3' end.

Gene 2 was cloned as a HindIII blunt ended fragment into the EcoRV site of pBluescript II SK+ vector from Stratagene. A part of the 3' untranslated sequence was removed by a SacI digestion, followed by religation. HindIII and HpaI restriction sites were introduced immediately in front of the start ATG by digestion with HindIII and NarI and religation in the presence of the following annealed oligonucleotides

10 AGCTTGTTAACATGTATCCAACCCTCACCTTCGTGG
ACAATTGTACATAGGTTGGGAGTGGAAGCACCGC

No introns were found in the clones sequenced.

15 The clone 1 type (SEQ.ID.No.3) can be aligned with all ten peptide sequences (see Figure 8) showing 100% identity. Alignment of the two protein sequences encoded by the genes isolated from the fungal infected algae *Gracilariopsis lemaneiformis* shows about 78% identity, indicating that both genes are coding for a α -1.4-glucan lyase.

20

5. EXPRESSION OF THE GL GENE IN MICRO-ORGANISMS

(E.G. ANALYSES OF *PICHIA* LYASE TRANSFORMANTS AND *ASPERGILLUS* LYASE TRANSFORMANTS)

25 The DNA sequence encoding the GL was introduced into microorganisms to produce an enzyme with high specific activity and in large quantities.

30 In this regard, gene 1 (i.e. SEQ. I.D. No. 3) was cloned as a NotI-HindIII blunt ended (using the DNA blunting kit from Amersham International) fragment into the *Pichia* expression vector pHIL-D2 (containing the AOX1 promoter) digested with EcoRI and blunt ended (using the DNA blunting kit from Amersham International) for expression in *Pichia pastoris* (according to the protocol stated in the *Pichia*

Expression Kit supplied by Invitrogen).

In another embodiment, the gene 1 (i.e. SEQ. I.D. No. 3) was cloned as a NotI-HindIII blunt ended fragment (using the DNA blunting kit from Amersham International) into the *Aspergillus* expression vector pBARMTE1 (containing the methyl tryptophan resistance promoter from *Neurospora crassa*) digested with SmaI for expression in *Aspergillus niger* (Pall et al (1993) Fungal Genet Newslett. vol 40 pages 59-62). The protoplasts were prepared according to Daboussi et al (Curr Genet (1989) vol 15 pp 453-456) using lysing enzymes Sigma L-2773 and the lyticase Sigma L-8012. The transformation of the protoplasts was followed according to the protocol stated by Buxton et al (Gene (1985) vol 37 pp 207-214) except that for plating the transformed protoplasts the protocol laid out in Punt et al (Methods in Enzymology (1992) vol 216 pp 447 - 457) was followed but with the use of 0.6% osmotic stabilised top agarose.

The results showed that lyase activity was observed in the transformed *Pichia pastoris* and *Aspergillus niger*.

5.1 GENERAL METHODS

Preparation of cell-free extracts.

The cells were harvested by centrifugation at 9000 rpm for 5 min and washed with 0.9% NaCl and resuspended in the breaking buffer (50mM K-phosphate, pH 7.5 containing 1mM of EDTA, and 5% glycerol). Cells were broken using glass beads and vortex treatment. The breaking buffer contained 1 mM PMSF (protease inhibitor). The lyase extract (supernatant) was obtained after centrifugation at 9000 rpm for 5 min followed by centrifugation at 20,000 xg for 5min.

Assay of lyase activity by alkaline 3,5-dinitrosalicylic acid reagent (DNS)

One volume of lyase extract was mixed with an equal volume of 4% amylopectin

solution. The reaction mixture was then incubated at a controlled temperature and samples were removed at specified intervals and analyzed for AF.

The lyase activity was also analyzed using a radioactive method.

5

The reaction mixture contained 10 μ l 14 C-starch solution (1 μ Ci; Sigma Chemicals Co.) and 10 μ l of the lyase extract. The reaction mixture was left at 25°C overnight and was then analyzed in the usual TLC system. The radioactive AF produced was detected using an Instant Imager (Pachard Instrument Co., Inc., Meriden, CT).

10

Electrophoresis and Western blotting

SDS-PAGE was performed using 8-25% gradient gels and the PhastSystem (Pharmacia). Western blottings was also run on a Semidry transfer unit of the PhastSystem.

15

Primary antibodies raised against the lyase purified from the red seaweed collected at Qingdao (China) were used in a dilution of 1:100. Pig antirabbit IgG conjugated to alkaline phosphatase (Dako A/S, Glostrup, Denmark) were used as secondary antibodies and used in a dilution of 1:1000.

20

Part I, Analysis of the Pichia transformants containing the above mentioned construct

25

Results:

1. Lyase activity was determined 5 days after induction (according to the manual) and proved the activity to be intracellular for all samples in the B series.

Samples of B series: 11 12 13 15 26 27 28 29 30

Specific activity: 139 81 122 192 151 253 199 198 150

*Specific activity is defined as nmol AF released per min per mg protein in a reaction mixture containing 2% (w/v) of glycogen, 1% (w/v) glycerol in 10 mM potassium phosphate buffer (pH 7.5). The reaction temperature was 45°C; the reaction time was 60 min.

A time course of sample B27 is as follows. The data are also presented in Figure 1.

Time (min) 0 10 20 30 40 50 60

Spec. act. 0 18 54 90 147 179 253

Assay conditions were as above except that the time was varied.

2. Western-blotting analysis.

The CFE of all samples showed bands with a molecular weight corresponding to the native lyase.

 MC-Lyase expressed intracellularly in *Pichia pastoris*

5	Names of culture	Specific activity*
	A18	10
	A20	32
10	A21	8
	A22	8
	A24	6
15		

Part II, The *Aspergillus* transformants

Results

20

I. Lyase activity was determined after 5 days incubation(minimal medium containing 0.2% casein enzymatic hydrolysate analysis by the alkaline 3,5-dinitrosalicylic acid reagent

25

1). Lyase activity analysis of the culture medium

30

Among 35 cultures grown with 0.2% amylopectin included in the culture medium, AF was only detectable in two cultures. The culture medium of 5.4+ and 5.9+ contained 0.13 g AF/liter and 0.44 g/liter, respectively. The result indicated that active lyase had been secreted from the cells. Lyase activity was also measurable in the cell-free extract.

2). Lyase activity analysis in cell-free extracts

	Name of the culture	Specific activity*
5	5.4+	51
	5.9+	148
10	5.13	99
	5.15	25
	5.19	37
15		

*The specific activity was defined as nmol of AF produced per min per mg protein at 25°C. + indicates that 0.2% amylopectin was added.

The results show that Gene 1 of GL was expressed intracellular in *A. niger*.

Experiments with transformed *E. coli* (using cloning vectors pQE30 from the Qia express vector kit from Qiagen) showed expression of enzyme that was recognised by anti-body to the enzyme purified from fungally infected *Gracilariopsis lemaneiformis*.

Instead of *Aspergillus niger* as host, other industrial important microorganisms for which good expression systems are known could be used such as: *Aspergillus oryzae*, *Aspergillus sp.*, *Trichoderma sp.*, *Saccharomyces cerevisiae*, *Kluyveromyces sp.*, *Hansenula sp.*, *Pichia sp.*, *Bacillus subtilis*, *B. amyloliquefaciens*, *Bacillus sp.*, *Streptomyces sp.* or *E. coli*.

Other preferred embodiments of the present invention include any one of the following: A transformed host organism having the capability of producing AF as a consequence of the introduction of a DNA sequence as herein described; such a transformed host organism which is a microorganism - preferably wherein the host organism is selected from the group consisting of bacteria, moulds, fungi and yeast; preferably the host organism is selected from the group consisting of *Saccharomyces*, *Kluyveromyces*, *Aspergillus*, *Trichoderma Hansenula*, *Pichia*, *Bacillus Streptomyces*, *Eschericia* such as *Aspergillus oryzae*, *Saccharomyces cerevisiae*, *bacillus subtilis*, *Bacillus amyloliquefascien*, *Eschericia coli*.; A method for preparing the sugar 1,5-D-anhydrofructose comprising contacting an alpha 1,4-glucan (e.g. starch) with the enzyme α -1,4-glucan lyase expressed by a transformed host organism comprising a nucleotide sequence encoding the same, preferably wherein the nucleotide sequence is a DNA sequence, preferably wherein the DNA sequence is one of the sequences hereinbefore described; A vector incorporating a nucleotide sequence as hereinbefore described, preferably wherein the vector is a replication vector, preferably wherein the vector is an expression vector containing the nucleotide sequence downstream from a promoter sequence, the vector preferably containing a marker (such as a resistance marker); Cellular organisms, or cell line, transformed with such a vector; A method of producing the product α -1,4-glucan lyase or any nucleotide sequence or part thereof coding for same, which comprises culturing such an organism (or cells from a cell line) transfected with such a vector and recovering the product.

Other modifications of the present invention will be apparent to those skilled in the art without departing from the scope of the invention.

SEQUENCE LISTING

(1) GENERAL INFORMATION:

(i) APPLICANT:

- (A) NAME: DANISCO A/S
- (B) STREET: LANGEBROGADE 1
- (C) CITY: COPENHAGEN
- (D) STATE: COPENHAGEN K
- (E) COUNTRY: DENMARK
- (F) POSTAL CODE (ZIP): DK-1001

(ii) TITLE OF INVENTION: ENZYME

(iii) NUMBER OF SEQUENCES: 20

(iv) COMPUTER READABLE FORM:

- (A) MEDIUM TYPE: Floppy disk
- (B) COMPUTER: IBM PC compatible
- (C) OPERATING SYSTEM: PC-DOS/MS-DOS
- (D) SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)

(v) CURRENT APPLICATION DATA:

APPLICATION NUMBER: WO PCT/EP94/03399

(2) INFORMATION FOR SEQ ID NO: 1:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1088 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 1:

```

Met Phe Ser Thr Leu Ala Phe Val Ala Pro Ser Ala Leu Gly Ala Ser
1           5           10           15

Thr Phe Val Gly Ala Glu Val Arg Ser Asn Val Arg Ile His Ser Ala
          20           25           30

Phe Pro Ala Val His Thr Ala Thr Arg Lys Thr Asn Arg Leu Asn Val
          35           40           45

Ser Met Thr Ala Leu Ser Asp Lys Gln Thr Ala Thr Ala Gly Ser Thr
          50           55           60

Asp Asn Pro Asp Gly Ile Asp Tyr Lys Thr Tyr Asp Tyr Val Gly Val
65           70           75           80

Trp Gly Phe Ser Pro Leu Ser Asn Thr Asn Trp Phe Ala Ala Gly Ser
          85           90           95

```

31

Ser Thr Pro Gly Gly Ile Thr Asp Trp Thr Ala Thr Met Asn Val Asn
 100 105 110
 Phe Asp Arg Ile Asp Asn Pro Ser Ile Thr Val Gln His Pro Val Gln
 115 120 125
 Val Gln Val Thr Ser Tyr Asn Asn Asn Ser Tyr Arg Val Arg Phe Asn
 130 135 140
 Pro Asp Gly Pro Ile Arg Asp Val Thr Arg Gly Pro Ile Leu Lys Gln
 145 150 155 160
 Gln Leu Asp Trp Ile Arg Thr Gln Glu Leu Ser Glu Gly Cys Asp Pro
 165 170 175
 Gly Met Thr Phe Thr Ser Glu Gly Phe Leu Thr Phe Glu Thr Lys Asp
 180 185 190
 Leu Ser Val Ile Ile Tyr Gly Asn Phe Lys Thr Arg Val Thr Arg Lys
 195 200 205
 Ser Asp Gly Lys Val Ile Met Glu Asn Asp Glu Val Gly Thr Ala Ser
 210 215 220
 Ser Gly Asn Lys Cys Arg Gly Leu Met Phe Val Asp Arg Leu Tyr Gly
 225 230 235 240
 Asn Ala Ile Ala Ser Val Asn Lys Asn Phe Arg Asn Asp Ala Val Lys
 245 250 255
 Gln Glu Gly Phe Tyr Gly Ala Gly Glu Val Asn Cys Lys Tyr Gln Asp
 260 265 270
 Thr Tyr Ile Leu Glu Arg Thr Gly Ile Ala Met Thr Asn Tyr Asn Tyr
 275 280 285
 Asp Asn Leu Asn Tyr Asn Gln Trp Asp Leu Arg Pro Pro His His Asp
 290 295 300
 Gly Ala Leu Asn Pro Asp Tyr Tyr Ile Pro Met Tyr Tyr Ala Ala Pro
 305 310 315 320
 Trp Leu Ile Val Asn Gly Cys Ala Gly Thr Ser Glu Gln Tyr Ser Tyr
 325 330 335
 Gly Trp Phe Met Asp Asn Val Ser Gln Ser Tyr Met Asn Thr Gly Asp
 340 345 350
 Thr Thr Trp Asn Ser Gly Gln Glu Asp Leu Ala Tyr Met Gly Ala Gln
 355 360 365
 Tyr Gly Pro Phe Asp Gln His Phe Val Tyr Gly Ala Gly Gly Gly Met
 370 375 380
 Glu Cys Val Val Thr Ala Phe Ser Leu Leu Gln Gly Lys Glu Phe Glu
 385 390 395 400

Asn	Gln	Val	Leu	Asn	Lys	Arg	Ser	Val	Met	Pro	Pro	Lys	Tyr	Val	Phe
				405					410					415	
Gly	Phe	Phe	Gln	Gly	Val	Phe	Gly	Thr	Ser	Ser	Leu	Leu	Arg	Ala	His
			420					425					430		
Met	Pro	Ala	Gly	Glu	Asn	Asn	Ile	Ser	Val	Glu	Glu	Ile	Val	Glu	Gly
		435					440					445			
Tyr	Gln	Asn	Asn	Asn	Phe	Pro	Phe	Glu	Gly	Leu	Ala	Val	Asp	Val	Asp
	450					455					460				
Met	Gln	Asp	Asn	Leu	Arg	Val	Phe	Thr	Thr	Lys	Gly	Glu	Phe	Trp	Thr
465					470					475					480
Ala	Asn	Arg	Val	Gly	Thr	Gly	Gly	Asp	Pro	Asn	Asn	Arg	Ser	Val	Phe
				485					490					495	
Glu	Trp	Ala	His	Asp	Lys	Gly	Leu	Val	Cys	Gln	Thr	Asn	Ile	Thr	Cys
			500					505					510		
Phe	Leu	Arg	Asn	Asp	Asn	Glu	Gly	Gln	Asp	Tyr	Glu	Val	Asn	Gln	Thr
		515					520					525			
Leu	Arg	Glu	Arg	Gln	Leu	Tyr	Thr	Lys	Asn	Asp	Ser	Leu	Thr	Gly	Thr
	530					535					540				
Asp	Phe	Gly	Met	Thr	Asp	Asp	Gly	Pro	Ser	Asp	Ala	Tyr	Ile	Gly	His
545					550					555					560
Leu	Asp	Tyr	Gly	Gly	Gly	Val	Glu	Cys	Asp	Ala	Leu	Phe	Pro	Asp	Trp
				565					570					575	
Gly	Arg	Pro	Asp	Val	Ala	Glu	Trp	Trp	Gly	Asn	Asn	Tyr	Lys	Lys	Leu
			580					585					590		
Phe	Ser	Ile	Gly	Leu	Asp	Phe	Val	Trp	Gln	Asp	Met	Thr	Val	Pro	Ala
		595					600					605			
Met	Met	Pro	His	Lys	Ile	Gly	Asp	Asp	Ile	Asn	Val	Lys	Pro	Asp	Gly
	610					615					620				
Asn	Trp	Pro	Asn	Ala	Asp	Asp	Pro	Ser	Asn	Gly	Gln	Tyr	Asn	Trp	Lys
625					630					635					640
Thr	Tyr	His	Pro	Gln	Val	Leu	Val	Thr	Asp	Met	Arg	Tyr	Glu	Asn	His
				645					650					655	
Gly	Arg	Glu	Pro	Met	Val	Thr	Gln	Arg	Asn	Ile	His	Ala	Tyr	Thr	Leu
			660					665					670		
Cys	Glu	Ser	Thr	Arg	Lys	Glu	Gly	Ile	Val	Glu	Asn	Ala	Asp	Thr	Leu
		675					680					685			
Thr	Lys	Phe	Arg	Arg	Ser	Tyr	Ile	Ile	Ser	Arg	Gly	Gly	Tyr	Ile	Gly
	690					695					700				

33

Asn 705	Gln	His	Phe	Gly	Gly 710	Met	Trp	Val	Gly	Asp 715	Asn	Ser	Thr	Thr	Ser 720
Asn	Tyr	Ile	Gln	Met 725	Met	Ile	Ala	Asn	Asn 730	Ile	Asn	Met	Asn	Met	Ser 735
Cys	Leu	Pro	Leu 740	Val	Gly	Ser	Asp	Ile 745	Gly	Gly	Phe	Thr	Ser 750	Tyr	Asp
Asn	Glu	Asn 755	Gln	Arg	Thr	Pro	Cys 760	Thr	Gly	Asp	Leu	Met 765	Val	Arg	Tyr
Val 770	Gln	Ala	Gly	Cys	Leu	Leu 775	Pro	Trp	Phe	Arg	Asn 780	His	Tyr	Asp	Arg
Trp 785	Ile	Glu	Ser	Lys	Asp 790	His	Gly	Lys	Asp	Tyr 795	Gln	Glu	Leu	Tyr	Met 800
Tyr	Pro	Asn	Glu	Met 805	Asp	Thr	Leu	Arg	Lys 810	Phe	Val	Glu	Phe	Arg 815	Tyr
Arg	Trp	Gln	Glu 820	Val	Leu	Tyr	Thr	Ala 825	Met	Tyr	Gln	Asn	Ala 830	Ala	Phe
Gly	Lys	Pro 835	Ile	Ile	Lys	Ala	Ala 840	Ser	Met	Tyr	Asn	Asn 845	Asp	Ser	Asn
Val 850	Arg	Arg	Ala	Gln	Asn	Asp 855	His	Phe	Leu	Leu	Gly 860	Gly	His	Asp	Gly
Tyr 865	Arg	Ile	Leu	Cys	Ala 870	Pro	Val	Val	Trp	Glu 875	Asn	Ser	Thr	Glu	Arg 880
Glu	Leu	Tyr	Leu	Pro 885	Val	Leu	Thr	Gln	Trp 890	Tyr	Lys	Phe	Gly	Pro 895	Asp
Phe	Asp	Thr	Lys 900	Pro	Leu	Glu	Gly	Ala 905	Met	Asn	Gly	Gly	Asp 910	Arg	Ile
Tyr	Asn	Tyr 915	Pro	Val	Pro	Gln	Ser 920	Glu	Ser	Pro	Ile	Phe 925	Val	Arg	Glu
Gly 930	Ala	Ile	Leu	Pro	Thr	Arg 935	Tyr	Thr	Leu	Asn	Gly 940	Glu	Asn	Lys	Ser
Leu 945	Asn	Thr	Tyr	Thr	Asp 950	Glu	Asp	Pro	Leu	Val 955	Phe	Glu	Val	Phe	Pro 960
Leu	Gly	Asn	Asn	Arg 965	Ala	Asp	Gly	Met	Cys 970	Tyr	Leu	Asp	Asp	Gly 975	Gly
Val	Thr	Thr	Asn 980	Ala	Glu	Asp	Asn	Gly 985	Lys	Phe	Ser	Val	Val 990	Lys	Val
Ala	Ala	Glu	Gln	Asp	Gly	Gly	Thr	Glu	Thr	Ile	Thr	Phe	Thr	Asn	Asp 1005

SUBSTITUTE SHEET (RULE 26)

34

Cys Tyr Glu Tyr Val Phe Gly Gly Pro Phe Tyr Val Arg Val Arg Gly
 1010 1015 1020
 Ala Gln Ser Pro Ser Asn Ile His Val Ser Ser Gly Ala Gly Ser Gln
 1025 1030 1035 1040
 Asp Met Lys Val Ser Ser Ala Thr Ser Arg Ala Ala Leu Phe Asn Asp
 1045 1050 1055
 Gly Glu Asn Gly Asp Phe Trp Val Asp Gln Glu Thr Asp Ser Leu Trp
 1060 1065 1070
 Leu Lys Leu Pro Asn Val Val Leu Pro Asp Ala Val Ile Thr Ile Thr
 1075 1080 1085

(2) INFORMATION FOR SEQ ID NO: 2:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1091 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 2:

Met Tyr Pro Thr Leu Thr Phe Val Ala Pro Ser Ala Leu Gly Ala Arg
 1 5 10 15
 Thr Phe Thr Cys Val Gly Ile Phe Arg Ser His Ile Leu Ile His Ser
 20 25 30
 Val Val Pro Ala Val Arg Leu Ala Val Arg Lys Ser Asn Arg Leu Asn
 35 40 45
 Val Ser Met Ser Ala Leu Phe Asp Lys Pro Thr Ala Val Thr Gly Gly
 50 55 60
 Lys Asp Asn Pro Asp Asn Ile Asn Tyr Thr Thr Tyr Asp Tyr Val Pro
 65 70 75 80
 Val Trp Arg Phe Asp Pro Leu Ser Asn Thr Asn Trp Phe Ala Ala Gly
 85 90 95
 Ser Ser Thr Pro Gly Asp Ile Asp Asp Trp Thr Ala Thr Met Asn Val
 100 105 110
 Asn Phe Asp Arg Ile Asp Asn Pro Ser Phe Thr Leu Glu Lys Pro Val
 115 120 125
 Gln Val Gln Val Thr Ser Tyr Lys Asn Asn Cys Phe Arg Val Arg Phe
 130 135 140

35

Asn	Pro	Asp	Gly	Pro	Ile	Arg	Asp	Val	Asp	Arg	Gly	Pro	Ile	Leu	Gln
145					150					155					160
Gln	Gln	Leu	Asn	Trp	Ile	Arg	Lys	Gln	Glu	Gln	Ser	Lys	Gly	Phe	Asp
			165						170					175	
Pro	Lys	Met	Gly	Phe	Thr	Lys	Glu	Gly	Phe	Leu	Lys	Phe	Glu	Thr	Lys
			180					185					190		
Asp	Leu	Asn	Val	Ile	Ile	Tyr	Gly	Asn	Phe	Lys	Thr	Arg	Val	Thr	Arg
		195					200					205			
Lys	Arg	Asp	Gly	Lys	Gly	Ile	Met	Glu	Asn	Asn	Glu	Val	Pro	Ala	Gly
	210					215					220				
Ser	Leu	Gly	Asn	Lys	Cys	Arg	Gly	Leu	Met	Phe	Val	Asp	Arg	Leu	Tyr
225					230					235					240
Gly	Thr	Ala	Ile	Ala	Ser	Val	Asn	Glu	Asn	Tyr	Arg	Asn	Asp	Pro	Asp
				245					250					255	
Arg	Lys	Glu	Gly	Phe	Tyr	Gly	Ala	Gly	Glu	Val	Asn	Cys	Glu	Phe	Trp
			260					265					270		
Asp	Ser	Glu	Gln	Asn	Arg	Asn	Lys	Tyr	Ile	Leu	Glu	Arg	Thr	Gly	Ile
		275					280					285			
Ala	Met	Thr	Asn	Tyr	Asn	Tyr	Asp	Asn	Tyr	Asn	Tyr	Asn	Gln	Ser	Asp
	290					295					300				
Leu	Ile	Ala	Pro	Gly	Tyr	Pro	Ser	Asp	Pro	Asn	Phe	Tyr	Ile	Pro	Met
305					310					315					320
Tyr	Phe	Ala	Ala	Pro	Trp	Val	Val	Val	Lys	Gly	Cys	Ser	Gly	Asn	Ser
				325					330					335	
Asp	Glu	Gln	Tyr	Ser	Tyr	Gly	Trp	Phe	Met	Asp	Asn	Val	Ser	Gln	Thr
			340					345					350		
Tyr	Met	Asn	Thr	Gly	Gly	Thr	Ser	Trp	Asn	Cys	Gly	Glu	Glu	Asn	Leu
		355					360					365			
Ala	Tyr	Met	Gly	Ala	Gln	Cys	Gly	Pro	Phe	Asp	Gln	His	Phe	Val	Tyr
	370					375					380				
Gly	Asp	Gly	Asp	Gly	Leu	Glu	Asp	Val	Val	Gln	Ala	Phe	Ser	Leu	Leu
385					390					395					400
Gln	Gly	Lys	Glu	Phe	Glu	Asn	Gln	Val	Leu	Asn	Lys	Arg	Ala	Val	Met
				405					410					415	
Pro	Pro	Lys	Tyr	Val	Phe	Gly	Tyr	Phe	Gln	Gly	Val	Phe	Gly	Ile	Ala
			420					425					430		
Ser	Leu	Leu	Arg	Glu	Gln	Arg	Pro	Glu	Gly	Gly	Asn	Asn	Ile	Ser	Val
		435					440					445			

SUBSTITUTE SHEET (RULE 26)

Gln Glu Ile Val Glu Gly Tyr Gln Ser Asn Asn Phe Pro Leu Glu Gly
 450 455 460
 Leu Ala Val Asp Val Asp Met Gln Gln Asp Leu Arg Val Phe Thr Thr
 465 470 475 480
 Lys Ile Glu Phe Trp Thr Ala Asn Lys Val Gly Thr Gly Gly Asp Ser
 485 490 495
 Asn Asn Lys Ser Val Phe Glu Trp Ala His Asp Lys Gly Leu Val Cys
 500 505 510
 Gln Thr Asn Val Thr Cys Phe Leu Arg Asn Asp Asn Gly Gly Ala Asp
 515 520 525
 Tyr Glu Val Asn Gln Thr Leu Arg Glu Lys Gly Leu Tyr Thr Lys Asn
 530 535 540
 Asp Ser Leu Thr Asn Thr Asn Phe Gly Thr Thr Asn Asp Gly Pro Ser
 545 550 555 560
 Asp Ala Tyr Ile Gly His Leu Asp Tyr Gly Gly Gly Gly Asn Cys Asp
 565 570 575
 Ala Leu Phe Pro Asp Trp Gly Arg Pro Gly Val Ala Glu Trp Trp Gly
 580 585 590
 Asp Asn Tyr Ser Lys Leu Phe Lys Ile Gly Leu Asp Phe Val Trp Gln
 595 600 605
 Asp Met Thr Val Pro Ala Met Met Pro His Lys Val Gly Asp Ala Val
 610 615 620
 Asp Thr Arg Ser Pro Tyr Gly Trp Pro Asn Glu Asn Asp Pro Ser Asn
 625 630 635 640
 Gly Arg Tyr Asn Trp Lys Ser Tyr His Pro Gln Val Leu Val Thr Asp
 645 650 655
 Met Arg Tyr Glu Asn His Gly Arg Glu Pro Met Phe Thr Gln Arg Asn
 660 665 670
 Met His Ala Tyr Thr Leu Cys Glu Ser Thr Arg Lys Glu Gly Ile Val
 675 680 685
 Ala Asn Ala Asp Thr Leu Thr Lys Phe Arg Arg Ser Tyr Ile Ile Ser
 690 695 700
 Arg Gly Gly Tyr Ile Gly Asn Gln His Phe Gly Gly Met Trp Val Gly
 705 710 715 720
 Asp Asn Ser Ser Ser Gln Arg Tyr Leu Gln Met Met Ile Ala Asn Ile
 725 730 735
 Val Asn Met Asn Met Ser Cys Leu Pro Leu Val Gly Ser Asp Ile Gly
 740 745 750

37

Gly Phe Thr Ser Tyr Asp Gly Arg Asn Val Cys Pro Gly Asp Leu Met
 755 760 765
 Val Arg Phe Val Gln Ala Gly Cys Leu Leu Pro Trp Phe Arg Asn His
 770 775 780
 Tyr Gly Arg Leu Val Glu Gly Lys Gln Glu Gly Lys Tyr Tyr Gln Glu
 785 790 795 800
 Leu Tyr Met Tyr Lys Asp Glu Met Ala Thr Leu Arg Lys Phe Ile Glu
 805 810 815
 Phe Arg Tyr Arg Trp Gln Glu Val Leu Tyr Thr Ala Met Tyr Gln Asn
 820 825 830
 Ala Ala Phe Gly Lys Pro Ile Ile Lys Ala Ala Ser Met Tyr Asp Asn
 835 840 845
 Asp Arg Asn Val Arg Gly Ala Gln Asp Asp His Phe Leu Leu Gly Gly
 850 855 860
 His Asp Gly Tyr Arg Ile Leu Cys Ala Pro Val Val Trp Glu Asn Thr
 865 870 875 880
 Thr Ser Arg Asp Leu Tyr Leu Pro Val Leu Thr Lys Trp Tyr Lys Phe
 885 890 895
 Gly Pro Asp Tyr Asp Thr Lys Arg Leu Asp Ser Ala Leu Asp Gly Gly
 900 905 910
 Gln Met Ile Lys Asn Tyr Ser Val Pro Gln Ser Asp Ser Pro Ile Phe
 915 920 925
 Val Arg Glu Gly Ala Ile Leu Pro Thr Arg Tyr Thr Leu Asp Gly Ser
 930 935 940
 Asn Lys Ser Met Asn Thr Tyr Thr Asp Lys Asp Pro Leu Val Phe Glu
 945 950 955 960
 Val Phe Pro Leu Gly Asn Asn Arg Ala Asp Gly Met Cys Tyr Leu Asp
 965 970 975
 Asp Gly Gly Ile Thr Thr Asp Ala Glu Asp His Gly Lys Phe Ser Val
 980 985 990
 Ile Asn Val Glu Ala Leu Arg Lys Gly Val Thr Thr Thr Ile Lys Phe
 995 1000 1005
 Ala Tyr Asp Thr Tyr Gln Tyr Val Phe Asp Gly Pro Phe Tyr Val Arg
 1010 1015 1020
 Ile Arg Asn Leu Thr Thr Ala Ser Lys Ile Asn Val Ser Ser Gly Ala
 1025 1030 1035 1040
 Gly Glu Glu Asp Met Thr Pro Thr Ser Ala Asn Ser Arg Ala Ala Leu
 1045 1050 1055

SUBSTITUTE SHEET (RULE 26)

37 / 1

Phe Ser Asp Gly Gly Val Gly Glu Tyr Trp Ala Asp Asn Asp Thr Ser
 1060 1065 1070

Ser Leu Trp Met Lys Leu Pro Asn Leu Val Leu Gln Asp Ala Val Ile
 1075 1080 1085

Thr Ile Thr
 1090

(2) INFORMATION FOR SEQ ID NO: 3:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 3267 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 3:

ATGTTTTCAA CCCTTGC GTT TGTCGCACCT AGTGCCTGG GAGCCAGTAC CTTCGTAGGG	60
GCGGAGGTCA GGTCAAATGT TCGTATCCAT TCCGCTTTTC CAGCTGTGCA CACAGCTACT	120
CGCAAAACCA ATCGCCTCAA TGTATCCATG ACCGCATTGT CCGACAAACA AACGGCTACT	180
GCGGGTAGTA CAGACAATCC GGACGGTATC GACTACAAGA CCTACGATTA CGTCGGAGTA	240
TGGGGTTTCA GCCCCCTCTC CAACACGAAC TGGTTTGCTG CCGGCTCTTC TACCCCGGGT	300
GGCATCACTG ATTGGACGGC TACAATGAAT GTCAACTTCG ACCGTATCGA CAATCCGTCC	360
ATCACTGTCC AGCATCCCGT TCAGGTTTCA GTCACGTCAT ACAACAACAA CAGCTACAGG	420
GTTGCTTCA ACCCTGATGG CCCTATTCTG TATGTGACTC GTGGGCCTAT CCTCAAGCAG	480
CAACTAGATT GGATTGGAAC GCAGGAGCTG TCAGAGGGAT GTGATCCCGG AATGACTTTC	540
ACATCAGAAG GTTCTTGAC TTTTGAGACC AAGGATCTAA GCGTCATCAT CTACGGAAAT	600
TTCAAGACCA GAGTTACGAG AAAGTCTGAC GGCAAGGTCA TCATGGAAAA TGATGAAGTT	660
GGAAGTGCAT CGTCCGGGAA CAAGTGCCGG GGATTGATGT TCGTTGATAG ATTATACGGT	720
AACGCTATCG CTTCCGTCAA CAAGAACTTC CGCAACGACG CGGTCAAGCA GGAGGGATTC	780
TATGGTGCAG GTGAAGTCAA CTGTAAGTAC CAGGACACCT ACATCTTAGA ACGCACTGGA	840
ATCGCCATGA CAAATTACAA CTACGATAAC TTGAACTATA ACCAGTGGGA CCTTAGACCT	900
CCGCATCATG ATGGTGCCCT CAACCCAGAC TATTATATTC CAATGTACTA CGCAGCACCT	960
TGGTTGATCG TTAATGGATG CGCCGGTACT TCGGAGCAGT ACTCGTATGG ATGGTTCATG	1020

GACAATGTCT CTCAATCTTA CATGAATACT GGAGATACTA CCTGGAATTC TGGACAAGAG	1080
GACCTGGCAT ACATGGGCGC GCAGTATGGA CCATTTGACC AACATTTTGT TTACGGTGCT	1140
GGGGGTGGGA TGGAAATGTGT GGTACAGCG TTCTCTCTTC TACAAGGCAA GGAGTTCGAG	1200
AACCAAGTTC TCAACAAACG TTCAGTAATG CCTCCGAAAT ACGTCTTTGG TTTCTTCCAG	1260
GGTGTTTTCG GGACTTCTTC CTTGTTGAGA GCGCATATGC CAGCAGGTGA GAACAACATC	1320
TCAGTCGAAG AAATTGTAGA AGGTTATCAA AACAACAATT TCCCTTTCGA GGGGCTCGCT	1380
GTGGACGTGG ATATGCAAGA CAACTTGCGG GTGTTACCA CGAAGGGCGA ATTTTGGACC	1440
GCAAACAGGG TGGGTACTGG CGGGGATCCA AACAACCGAT CGGTTTTTGA ATGGGCACAT	1500
GACAAAGGCC TTGTTTGTCA GACAAATATA ACTTGCTTCC TGAGGAATGA TAACGAGGGG	1560
CAAGACTACG AGGTCAATCA GACGTTAAGG GAGAGGCAGT TGTACACGAA GAACGACTCC	1620
CTGACGGGTA CGGATTTTGG AATGACCGAC GACGGCCCCA GCGATGCGTA CATCGGTCAT	1680
CTGGACTATG GGGGTGGAGT AGAATGTGAT GCACTTTTCC CAGACTGGGG ACGGCCTGAC	1740
GTGGCCGAAT GGTGGGGAAA TAACTATAAG AAACTGTTCA GCATTGGTCT CGACTTCGTC	1800
TGGCAAGACA TGA CTGTTCC AGCAATGATG CCGCACAAAA TTGGCGATGA CATCAATGTG	1860
AAACCGGATG GGAATTGGCC GAATGCGGAC GATCCGTCCA ATGGACAATA CAACTGGAAG	1920
ACGTACCATC CCCAAGTGCT TGTAAGTATG ATGCGTTATG AGAATCATGG TCGGGAACCG	1980
ATGGTCACTC AACGCAACAT TCATGCGTAT ACACTGTGCG AGTCTACTAG GAAGGAAGGG	2040
ATCGTGAAAA ACGCAGACAC TCTAACGAAG TTCCGCCGTA GCTACATTAT CAGTCGTGGT	2100
GGTTACATTG GTAACCAGCA TTTCGGGGGT ATGTGGGTGG GAGACAACTC TACTACATCA	2160
AACTACATCC AAATGATGAT TGCCAACAAT ATTAACATGA ATATGTCTTG CTTGCCTCTC	2220
GTCGGCTCCG ACATTGGAGG ATTCACCTCA TACGACAATG AGAATCAGCG AACGCCGTGT	2280
ACCGGGGACT TGATGGTGAG GTATGTGCAG GCGGGCTGCC TGTTGCCGTG GTTCAGGAAC	2340
CACTATGATA GGTGGATCGA GTCCAAGGAC CACGGAAAGG ACTACCAGGA GCTGTACATG	2400
TATCCGAATG AAATGGATAC GTTGAGGAAG TTCGTTGAAT TCCGTTATCG CTGGCAGGAA	2460
GTGTTGTACA CGGCCATGTA CCAGAATGCG GCTTTCGGAA AGCCGATTAT CAAGGCTGCT	2520
TCGATGTACA ATAACGACTC AAACGTTGCG AGGGCGCAGA ACGATCATTT CCTTCTTGGT	2580
GGACATGATG GATATCGCAT TCTGTGCGCG CCTGTTGTGT GGGAGAATTC GACCGAACGC	2640
GAATTGTACT TGCCCGTGCT GACCCAATGG TACAAATTCG GTCCCGACTT TGACACCAAG	2700

37 / 3

CCTCTGGAAG GAGCGATGAA CGGAGGGGAC CGAATTTACA ACTACCCTGT ACCGCAAAGT	2760
GAATCACCAA TCTTCGTGAG AGAAGGTGCG ATTCTCCCTA CCCGCTACAC GTTGAACGGT	2820
GAAAACAAAT CATTGAACAC GTACACGGAC GAAGATCCGT TGGTGTGTTGA AGTATTCCCC	2880
CTCGGAAACA ACCGTGCCGA CGGTATGTGT TATCTTGATG ATGGCGGTGT GACCACCAAT	2940
GCTGAAGACA ATGGCAAGTT CTCTGTCGTC AAGGTGGCAG CGGAGCAGGA TGGTGGTACG	3000
GAGACGATAA CGTTTACGAA TGATTGCTAT GAGTACGTTT TCGGTGGACC GTTCTACGTT	3060
CGAGTGCGCG GCGCTCAGTC GCCGTCGAAC ATCCACGTGT CTTCTGGAGC GGGTTCTCAG	3120
GACATGAAGG TGAGCTCTGC CACTTCCAGG GCTGCGCTGT TCAATGACGG GGAGAACGGT	3180
GATTTCTGGG TTGACCAGGA GACAGATTCT CTGTGGCTGA AGTTGCCCAA CGTTGTTCTC	3240
CCGGACGCTG TGATCACAAT TACCTAA	3267

(2) INFORMATION FOR SEQ ID NO: 4:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 3276 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 4:

ATGTATCCAA CCCTCACCTT CGTGGCGCCT AGTGCGCTAG GGGCCAGAAC TTTCACGTGT	60
GTGGGCATTT TTAGGTCACA CATTCTTATT CATTGCGTTG TTCCAGCGGT GCGTCTAGCT	120
GTGCGCAAAA GCAACCGCCT CAATGTATCC ATGTCCGCTT TGTTGACAA ACCGACTGCT	180
GTTACTGGAG GGAAGGACAA CCCGGACAAT ATCAATTACA CCACTTATGA CTACGTCCCT	240
GTGTGGCGCT TCGACCCCCT CAGCAATACG AACTGGTTTG CTGCCGGATC TTCCACTCCC	300
GGCGATATTG ACGACTGGAC GGCGACAATG AATGTGAACT TCGACCGTAT CGACAATCCA	360
TCCTTCACTC TCGAGAAACC GGTTCAGGT CAGGTCACGT CATAAAGAA CAATTGTTTC	420
AGGGTTCGCT TCAACCCTGA TGGTCCTATT CGCGATGTGG ATCGTGGGCC TATCCTCCAG	480
CAGCAACTAA ATTGGATCCG GAAGCAGGAG CAGTCGAAGG GGTTTGATCC TAAGATGGGC	540
TTCACAAAAG AAGGTTTCTT GAAATTTGAG ACCAAGGATC TGAACGTTAT CATATATGGC	600
AATTTTAAGA CTAGAGTTAC GAGGAAGAGG GATGGAAAAG GGATCATGGA GAATAATGAA	660

GTGCCGGCAG GATCGTTAGG GAACAAGTGC CGGGGATTGA TGTTTGTCGA CAGGTTGTAC	720
GGCACTGCCA TCGCTTCCGT TAATGAAAAT TACCGCAACG ATCCCGACAG GAAAGAGGGG	780
TTCTATGGTG CAGGAGAAGT AAACTGCGAG TTTTGGGACT CCGAACAAAA CAGGAACAAG	840
TACATCTTAG AACGAACTGG AATCGCCATG ACAAATTACA ATTATGACAA CTATAACTAC	900
AACCAGTCAG ATCTTATTGC TCCAGGATAT CCTTCCGACC CGAACTTCTA CATTCCCATG	960
TATTTTGCAG CACCTTGGGT AGTTGTAAAG GGATGCAGTG GCAACAGCGA TGAACAGTAC	1020
TCGTACGGAT GGTTTATGGA TAATGTCTCC CAACTTACA TGAATACTGG TGGTACTTCC	1080
TGGAAGTGTG GAGAGGAGAA CTTGGCATACT ATGGGAGCAC AGTGCGGTCC ATTTGACCAA	1140
CATTTTGTGT ATGGTGATGG AGATGGTCTT GAGGATGTTG TCCAAGCGTT CTCTCTTCTG	1200
CAAGGCAAAG AGTTTGAGAA CCAAGTTCTG AACAAACGTG CCGTAATGCC TCCGAAATAT	1260
GTGTTTGGTT ACTTTCAGGG AGTCTTTGGG ATTGCTTCTT TGTTGAGAGA GCAAAGACCA	1320
GAGGGTGGTA ATAACATCTC TGTTCAAGAG ATTGTGGAAG GTTACCAAAG CAATAACTTC	1380
CCTTTAGAGG GGTTAGCCGT AGATGTGGAT ATGCAACAAG ATTTGCGCGT GTTCACCACG	1440
AAGATTGAAT TTTGGACGGC AAATAAGGTA GGCACCGGGG GAGACTCGAA TAACAAGTCG	1500
GTGTTTGAAT GGGCACATGA CAAAGGCCTT GTATGTCAGA CGAATGTTAC TTGCTTCTTG	1560
AGAAACGACA ACGGCGGGGC AGATTACGAA GTCAATCAGA CATTGAGGGA GAAGGGTTTG	1620
TACACGAAGA ATGACTCACT GACGAACACT AACTTCGGAA CTACCAACGA CGGGCCGAGC	1680
GATGCGTACA TTGGACATCT GGACTATGGT GGCGGAGGGA ATTGTGATGC ACTTTTCCCA	1740
GACTGGGGTC GACCGGGTGT GGCTGAATGG TGGGGTGATA ACTACAGCAA GCTCTTCAAA	1800
ATTGGTCTGG ATTTCTGTCTG GCAAGACATG ACAGTTCCAG CTATGATGCC ACACAAAGTT	1860
GGCGACGCAG TCGATACGAG ATCACCTTAC GGCTGGCCGA ATGAGAATGA TCCTTCGAAC	1920
GGACGATACA ATTGGAAATC TTACCATCCA CAAGTTCTCG TAACTGATAT GCGATATGAG	1980
AATCATGGAA GGAACCGAT GTTCACTCAA CGCAATATGC ATGCGTACAC ACTCTGTGAA	2040
TCTACGAGGA AGGAAGGGAT TGTTGCAAAT GCAGACACTC TAACGAAGTT CCGCCGCAGT	2100
TATATTATCA GTCGTGGAGG TTACATTGGC AACCAGCATT TTGGAGGAAT GTGGGTTGGA	2160
GACAACTCTT CCTCCCAAAG ATACCTCCAA ATGATGATCG CGAACATCGT CAACATGAAC	2220
ATGTCTTGCC TTCCACTAGT TGGGTCCGAC ATTGGAGGTT TTAATTCTGTA TGATGGACGA	2280
AACGTGTGTC CCGGGGATCT AATGGTAAGA TTCGTGCAGG CGGGTTGCTT ACTACCGTGG	2340

3 7 / 5

```

TTCAGAAACC ACTATGGTAG GTTGGTCGAG GGCAAGCAAG AGGGAAAATA CTATCAAGAA 2400
CTGTACATGT ACAAGGACGA GATGGCTACA TTGAGAAAAT TCATTGAATT CCGTTACCGC 2460
TGGCAGGAGG TGTGTACAC TGCTATGTAC CAGAATGCGG CTTTCGGGAA ACCGATTATC 2520
AAGGCAGCTT CCATGTACGA CAACGACAGA AACGTTTCGCG GCGCACAGGA TGACCACTTC 2580
CTTCTCGGCG GACACGATGG ATATCGTATT TTGTGTGCAC CTGTTGTGTG GGAGAATACA 2640
ACCAGTCGCG ATCTGTACTT GCCTGTGCTG ACCAAATGGT ACAAATTCGG CCCTGACTAT 2700
GACACCAAGC GCCTGGATTC TCGTTGGAT GGAGGGCAGA TGATTAAGAA CTATTCTGTG 2760
CCACAAAGCG ACTCTCCGAT ATTTGTGAGG GAAGGAGCTA TTCTCCCTAC CCGCTACACG 2820
TTGGACGGTT CGAACAAGTC AATGAACACG TACACAGACA AAGACCCGTT GGTGTTTGAG 2880
GTATTCCTC TTGAAACAA CCGTGCCGAC GGTATGTGTT ATCTTGATGA TGGCGGTATT 2940
ACTACAGATG CTGAGGACCA TGGCAAATTC TCTGTTATCA ATGTCGAAGC CTTACGGAAA 3000
GGTGTACGA CGACGATCAA GTTTGCGTAT GACACTTATC AATACGTATT TGATGGTCCA 3060
TTCTACGTTG GAATCCGTAA TCTTACGACT GCATCAAAAA TTAACGTGTC TTCTGGAGCG 3120
GGTGAAGAGG ACATGACACC GACCTCTGCG AACTCGAGGG CAGCTTTGTT CAGTGATGGA 3180
GGTGTGGAG AATACTGGGC TGACAATGAT ACGTCTTCTC TGTGGATGAA GTTGCCAAAC 3240
CTGTTCTGC AAGACGCTGT GATTACCATT ACGTAG 3276

```

(2) INFORMATION FOR SEQ ID NO: 5:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 90 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 5:

```

Tyr Arg Trp Gln Glu Val Leu Tyr Thr Ala Met Tyr Gln Asn Ala Ala
1           5           10           15
Phe Gly Lys Pro Ile Ile Lys Ala Ala Ser Met Tyr Asn Asn Asp Ser
20           25           30
Asn Val Arg Arg Ala Gln Asn Asp His Phe Leu Leu Gly Gly His Asp
35           40           45

```

37 / 6

Gly Tyr Arg Ile Leu Cys Ala Pro Val Val Trp Glu Asn Ser Thr Glu
 50 55 60

Arg Glu Leu Tyr Leu Pro Val Leu Thr Gln Trp Tyr Lys Phe Gly Pro
 65 70 75 80

Asp Phe Asp Thr Lys Pro Leu Glu Gly Ala
 85 90

(2) INFORMATION FOR SEQ ID NO: 6:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 23 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(ix) FEATURE:

- (A) NAME/KEY: misc_difference
- (B) LOCATION: replace(6, "")
- (D) OTHER INFORMATION: /note= "N is T or C"

(ix) FEATURE:

- (A) NAME/KEY: misc_difference
- (B) LOCATION: replace(9, "")
- (D) OTHER INFORMATION: /note= "N is C or T"

(ix) FEATURE:

- (A) NAME/KEY: misc_difference
- (B) LOCATION: replace(12, "")
- (D) OTHER INFORMATION: /note= "N is C or T"

(ix) FEATURE:

- (A) NAME/KEY: misc_difference
- (B) LOCATION: replace(15, "")
- (D) OTHER INFORMATION: /note= "N is C or T"

(ix) FEATURE:

- (A) NAME/KEY: misc_difference
- (B) LOCATION: replace(18, "")
- (D) OTHER INFORMATION: /note= "N is G or A or T or C"

(ix) FEATURE:

- (A) NAME/KEY: misc_difference
- (B) LOCATION: replace(21, "")
- (D) OTHER INFORMATION: /note= "N is C or T"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 6:

ATGTANAANA ANGANTCNAA NGT

37 / 7

(2) INFORMATION FOR SEQ ID NO: 7:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 23 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(ix) FEATURE:

- (A) NAME/KEY: misc_difference
- (B) LOCATION: replace(6, "")
- (D) OTHER INFORMATION: /note= "N is T or C"

(ix) FEATURE:

- (A) NAME/KEY: misc_difference
- (B) LOCATION: replace(9, "")
- (D) OTHER INFORMATION: /note= "N is C or T"

(ix) FEATURE:

- (A) NAME/KEY: misc_difference
- (B) LOCATION: replace(12, "")
- (D) OTHER INFORMATION: /note= "N is C or T"

(ix) FEATURE:

- (A) NAME/KEY: misc_difference
- (B) LOCATION: replace(15, "")
- (D) OTHER INFORMATION: /note= "N is C or T"

(ix) FEATURE:

- (A) NAME/KEY: misc_difference
- (B) LOCATION: replace(18, "")
- (D) OTHER INFORMATION: /note= "N is C or T"

(ix) FEATURE:

- (A) NAME/KEY: misc_difference
- (B) LOCATION: replace(21, "")
- (D) OTHER INFORMATION: /note= "N is C or T"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 7:

ATGTANAANA ANGANAGNAA NGT

23

(2) INFORMATION FOR SEQ ID NO: 8:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 17 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

37 / 8

- (ix) FEATURE:
 - (A) NAME/KEY: misc difference
 - (B) LOCATION: replace(3, "")
 - (D) OTHER INFORMATION: /note= "N is G or A or T or C"
- (ix) FEATURE:
 - (A) NAME/KEY: misc difference
 - (B) LOCATION: replace(6, "")
 - (D) OTHER INFORMATION: /note= "N is G or A"
- (ix) FEATURE:
 - (A) NAME/KEY: misc difference
 - (B) LOCATION: replace(9, "")
 - (D) OTHER INFORMATION: /note= "N is G or A"
- (ix) FEATURE:
 - (A) NAME/KEY: misc difference
 - (B) LOCATION: replace(12, "")
 - (D) OTHER INFORMATION: /note= "N is G or A or T or C"
- (ix) FEATURE:
 - (A) NAME/KEY: misc difference
 - (B) LOCATION: replace(15, "")
 - (D) OTHER INFORMATION: /note= "N is G or A or T or C"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 8:

TANCCNTCNT GNCCNCC

17

(2) INFORMATION FOR SEQ ID NO: 9:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 20 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: DNA (genomic)
- (ix) FEATURE:
 - (A) NAME/KEY: misc difference
 - (B) LOCATION: replace(3, "")
 - (D) OTHER INFORMATION: /note= "N is G or A or T or C"
- (ix) FEATURE:
 - (A) NAME/KEY: misc difference
 - (B) LOCATION: replace(6, "")
 - (D) OTHER INFORMATION: /note= "N is G or A"
- (ix) FEATURE:
 - (A) NAME/KEY: misc difference
 - (B) LOCATION: replace(9, "")
 - (D) OTHER INFORMATION: /note= "N is C or T"

37 / 9

- (ix) FEATURE:
 (A) NAME/KEY: misc difference
 (B) LOCATION: replace(12, "")
 (D) OTHER INFORMATION: /note= "N is G or A"

- (ix) FEATURE:
 (A) NAME/KEY: misc difference
 (B) LOCATION: replace(18, "")
 (D) OTHER INFORMATION: /note= "N is C or T"

- (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 9:

GGNCCNAANT TNTACCANTG

20

- (2) INFORMATION FOR SEQ ID NO: 10:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 17 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

- (ii) MOLECULE TYPE: DNA (genomic)

- (ix) FEATURE:
 (A) NAME/KEY: misc difference
 (B) LOCATION: replace(3, "")
 (D) OTHER INFORMATION: /note= "N is T or C"

- (ix) FEATURE:
 (A) NAME/KEY: misc difference
 (B) LOCATION: replace(6, "")
 (D) OTHER INFORMATION: /note= "N is G or A or T or C"

- (ix) FEATURE:
 (A) NAME/KEY: misc difference
 (B) LOCATION: replace(12, "")
 (D) OTHER INFORMATION: /note= "N is G or A"

- (ix) FEATURE:
 (A) NAME/KEY: misc difference
 (B) LOCATION: replace(15, "")
 (D) OTHER INFORMATION: /note= "N is G or A"

- (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 10:

TANCGNTGGC ANGANGT

17

3 7 / 1 0

(2) INFORMATION FOR SEQ ID NO: 11:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 17 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(ix) FEATURE:

- (A) NAME/KEY: misc_difference
- (B) LOCATION: replace(3, "")
- (D) OTHER INFORMATION: /note= "N is T or C"

(ix) FEATURE:

- (A) NAME/KEY: misc_difference
- (B) LOCATION: replace(6, "")
- (D) OTHER INFORMATION: /note= "N is G or A"

(ix) FEATURE:

- (A) NAME/KEY: misc_difference
- (B) LOCATION: replace(12, "")
- (D) OTHER INFORMATION: /note= "N is G or A"

(ix) FEATURE:

- (A) NAME/KEY: misc_difference
- (B) LOCATION: replace(15, "")
- (D) OTHER INFORMATION: /note= "N is G or A"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 11:

TANAGNTGGC ANGANGT

17

(2) INFORMATION FOR SEQ ID NO: 12:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 71 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 12:

ATGTACAACA ACGACTCGAA CGTTCGCAGG GCGCAGAACG ATCATTTTCCT TCTTGCGCGC

60

CACGACGGTT A

71

37 / 11

(2) INFORMATION FOR SEQ ID NO: 13:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 23 amino acids
 (B) TYPE: amino acid
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 13:

Met	Tyr	Asn	Asn	Asp	Ser	Asn	Val	Arg	Arg	Ala	Gln	Asn	Asp	His	Phe
1				5				10						15	
Leu	Leu	Gly	Gly	His	Asp	Gly									
			20												

(2) INFORMATION FOR SEQ ID NO: 14:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 160 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 14:

ATGTACAACA	ACGACTCGAA	CGTTCGCAGG	GCGCAGAACG	ATCATTTTCCT	TCTTGGTGGA	60
CATGATGGAT	ATCGCATTCT	GTGCGCGCCT	GTTGTGTGGG	AGAATTCGAC	CGAACGGAAT	120
TGTACTTGCC	CGTGCTGACC	CAATGGTACA	AATTCGGCCC			160

(2) INFORMATION FOR SEQ ID NO: 15:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 54 amino acids
 (B) TYPE: amino acid
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 15:

Met	Tyr	Asn	Asn	Asp	Ser	Asn	Val	Arg	Arg	Ala	Gln	Asn	Asp	His	Phe
1				5				10						15	

37 / 12

Leu Leu Gly Gly His Asp Gly Tyr Arg Ile Leu Cys Ala Pro Val Val
20 25 30

Trp Glu Asn Ser Thr Glu Arg Glu Leu Tyr Leu Pro Val Leu Thr Gln
35 40 45

Trp Tyr Lys Phe Gly Pro
50

(2) INFORMATION FOR SEQ ID NO: 16:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 238 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 16:

TACAGGTGGC AGGAGGTGTT GTACACTGCT ATGTACCAGA ATGCGGCTTT CGGGAAACCG 60

ATTATCAAGG CAGCTTCCAT GTACGACAAC GACAGAAACG TTCGCGGCGC ACAGGATGAC 120

CACTTCCTTC TCGGCGGACA CGATGGATAT CGTATTTTGT GTGCACCTGT TGTGTGGGAG 180

AATACAACCA GTCGCGATCT GTACTTGCCT GTGCTGACCA GTGGTACAAA TTCGGCCC 238

(2) INFORMATION FOR SEQ ID NO: 17:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 79 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 17:

Tyr Arg Trp Gln Glu Val Leu Tyr Thr Ala Met Tyr Gln Asn Ala Ala
1 5 10 15

Phe Gly Lys Pro Ile Ile Lys Ala Ala Ser Met Tyr Asp Asn Asp Arg
20 25 30

Asn Val Arg Gly Ala Gln Asp Asp His Phe Leu Leu Gly Gly His Asp
35 40 45

Gly Tyr Arg Ile Leu Cys Ala Pro Val Val Trp Glu Asn Thr Thr Ser
50 55 60

3 7 / 1 3

Arg Asp Leu Tyr Leu Pro Val Leu Thr Lys Trp Tyr Lys Phe Gly
65 70 75

(2) INFORMATION FOR SEQ ID NO: 18:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 28 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 18:

GCTCTAGAGC ATGTTTTCAA CCCTTGCG

28

(2) INFORMATION FOR SEQ ID NO: 19:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 36 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 19:

AGCTTGTTAA CATGTATCCA ACCCTCACCT TCGTGG

36

(2) INFORMATION FOR SEQ ID NO: 20:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 34 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 20:

ACAATTGTAC ATAGGTTGGG AGTGGAAGCA CCGC

34

INDICATIONS RELATING TO A DEPOSITED MICROORGANISM

(PCT Rule 13bis)

A. The indications made below relate to the microorganism referred to in the description on page <u>5</u> , line <u>22</u>	
B. IDENTIFICATION OF DEPOSIT Further deposits are identified on an additional sheet <input type="checkbox"/>	
Name of depositary institution <p style="margin-left: 40px;">The National Collections of Industrial and Marine Bacteria Limited (NCIMB)</p>	
Address of depositary institution (including postal code and country) <p style="margin-left: 40px;">23 St. Machar Drive Aberdeen Scotland AB2 1RY United Kingdom</p>	
Date of deposit <p style="margin-left: 40px;">20 JUNE 1994</p>	Accession Number <p style="margin-left: 40px;">NCIMB 40652</p>
C. ADDITIONAL INDICATIONS (leave blank if not applicable) This information is continued on an additional sheet <input type="checkbox"/>	
<p>In respect of those designations in which a European patent is sought, and any other designated state having equivalent legislation, a sample of the deposited microorganism will be made available until the publication of the mention of the grant of the European patent or until the date on which the application has been refused or withdrawn or is deemed to be withdrawn, only by the issue of such a sample to an expert nominated by the person requesting the sample. (Rule 28(4) EPC).</p>	
D. DESIGNATED STATES FOR WHICH INDICATIONS ARE MADE (if the indications are not for all designated States)	
E. SEPARATE FURNISHING OF INDICATIONS (leave blank if not applicable)	
The indications listed below will be submitted to the International Bureau later (specify the general nature of the indications e.g., "Accession Number of Deposit")	

<p style="text-align: center;">For receiving Office use only</p> <div style="border: 1px solid black; padding: 5px; margin-bottom: 5px;"> <input type="checkbox"/> This sheet was received with the international application </div> <div style="border: 1px solid black; padding: 5px; min-height: 40px;"> Authorized officer </div>	<p style="text-align: center;">For International Bureau use only</p> <div style="border: 1px solid black; padding: 5px; margin-bottom: 5px;"> <input type="checkbox"/> This sheet was received by the International Bureau on: </div> <div style="border: 1px solid black; padding: 5px; min-height: 40px;"> Authorized officer </div>
---	--

INDICATIONS RELATING TO A DEPOSITED MICROORGANISM

(PCT Rule 13bis)

A. The indications made below relate to the microorganism referred to in the description on page <u>5</u> , line <u>24</u>	
B. IDENTIFICATION OF DEPOSIT Further deposits are identified on an additional sheet <input type="checkbox"/>	
Name of depositary institution <p style="margin-left: 40px;">The National Collections of Industrial and Marine Bacteria Limited (NCIMB)</p>	
Address of depositary institution (including postal code and country) <p style="margin-left: 40px;">23 St. Machar Drive Aberdeen Scotland AB2 1RY United Kingdom</p>	
Date of deposit <p style="margin-left: 40px;">20 JUNE 1994</p>	Accession Number <p style="margin-left: 40px;">NCIMB 40653</p>
C. ADDITIONAL INDICATIONS (leave blank if not applicable) This information is continued on an additional sheet <input type="checkbox"/>	
<p>In respect of those designations in which a European patent is sought, and any other designated state having equivalent legislation, a sample of the deposited microorganism will be made available until the publication of the mention of the grant of the European patent or until the date on which the application has been refused or withdrawn or is deemed to be withdrawn, only by the issue of such a sample to an expert nominated by the person requesting the sample. (Rule 28(4) EPC).</p>	
D. DESIGNATED STATES FOR WHICH INDICATIONS ARE MADE (if the indications are not for all designated States)	
E. SEPARATE FURNISHING OF INDICATIONS (leave blank if not applicable)	
The indications listed below will be submitted to the International Bureau later (specify the general nature of the indications e.g., "Accession Number of Deposit")	

<p style="text-align: center;">For receiving Office use only</p> <div style="border: 1px solid black; padding: 5px; margin-bottom: 5px;"> <input type="checkbox"/> This sheet was received with the international application </div> <div style="border: 1px solid black; padding: 5px; min-height: 40px;"> Authorized officer </div>	<p style="text-align: center;">For International Bureau use only</p> <div style="border: 1px solid black; padding: 5px; margin-bottom: 5px;"> <input type="checkbox"/> This sheet was received by the International Bureau on: </div> <div style="border: 1px solid black; padding: 5px; min-height: 40px;"> Authorized officer </div>
---	--

INDICATIONS RELATING TO A DEPOSITED MICROORGANISM

(PCT Rule 13bis)

A. The indications made below relate to the microorganism referred to in the description on page <u>5</u> , line <u>31</u>	
B. IDENTIFICATION OF DEPOSIT <div style="text-align: right;">Further deposits are identified on an additional sheet <input type="checkbox"/></div>	
Name of depositary institution Culture Collection of Algae and Protozoa (CCAP)	
Address of depositary institution (including postal code and country) Dunstaffnage Marine Laboratory P.O. Box 3 Oban Argyll PA34 4AD Scotland United Kingdom	
Date of deposit 11 OCTOBER 1994	Accession Number CCAP 1373/1
C. ADDITIONAL INDICATIONS (leave blank if not applicable) <div style="text-align: right;">This information is continued on an additional sheet <input type="checkbox"/></div>	
In respect of those designations in which a European patent is sought, and any other designated state having equivalent legislation, a sample of the deposited microorganism will be made available until the publication of the mention of the grant of the European patent or until the date on which the application has been refused or withdrawn or is deemed to be withdrawn, only by the issue of such a sample to an expert nominated by the person requesting the sample. (Rule 28(4) EPC).	
D. DESIGNATED STATES FOR WHICH INDICATIONS ARE MADE (if the indications are not for all designated States)	
E. SEPARATE FURNISHING OF INDICATIONS (leave blank if not applicable)	
The indications listed below will be submitted to the International Bureau later (specify the general nature of the indications e.g., "Accession Number of Deposit")	

<div style="text-align: center;">For receiving Office use only</div> <div style="border: 1px solid black; padding: 5px;"> <input type="checkbox"/> This sheet was received with the international application </div> <div style="border: 1px solid black; padding: 5px; margin-top: 10px;"> Authorized officer </div>	<div style="text-align: center;">For International Bureau use only</div> <div style="border: 1px solid black; padding: 5px;"> <input type="checkbox"/> This sheet was received by the International Bureau on: </div> <div style="border: 1px solid black; padding: 5px; margin-top: 10px;"> Authorized officer </div>
---	--

CLAIMS

1. A method of preparing the enzyme α -1,4-glucan lyase (GL) comprising isolating the enzyme from a fungally infected algae.
5
2. A method according to claim 1 wherein the enzyme is isolated and/or further purified using a gel that is not degraded by the enzyme.
3. A method according to claim 2 wherein the gel is based on dextrin or derivatives thereof, preferably a cyclodextrin, more preferably beta-cyclo-dextrin.
10
4. A GL enzyme prepared by the method according to any one of claims 1 to 3.
5. An enzyme comprising the amino acid sequence SEQ. ID. No. 1. or SEQ. ID. No. 2, or any variant thereof.
15
6. A nucleotide sequence coding for the enzyme α -1,4-glucan lyase.
7. A nucleotide sequence according to claim 6 wherein the sequence is a DNA sequence.
20
8. A nucleotide sequence according to claim 7 wherein the DNA sequence comprises a sequence that is the same as, or is complementary to, or has substantial homology with, or contains any suitable codon substitutions for any of those of, SEQ. ID. No. 3 or SEQ. ID. No. 4.
25
9. A method of preparing the enzyme α -1,4-glucan lyase comprising expressing the nucleotide sequence according to any one of claims 6 to 8.
30

10. A method according to any one of the preceding claims wherein the algae is *Gracilariopsis lemaneiformis*.

11. The use of beta-cyclodextrin to purify an enzyme, preferably GL.

5

12. A nucleotide sequence wherein the DNA sequence comprises a sequence that is the same as, or is complementary to, or has substantial homology with, or contains any suitable codon substitutions for any of those of, SEQ. ID. No. 3 or SEQ. ID. No. 4.

10

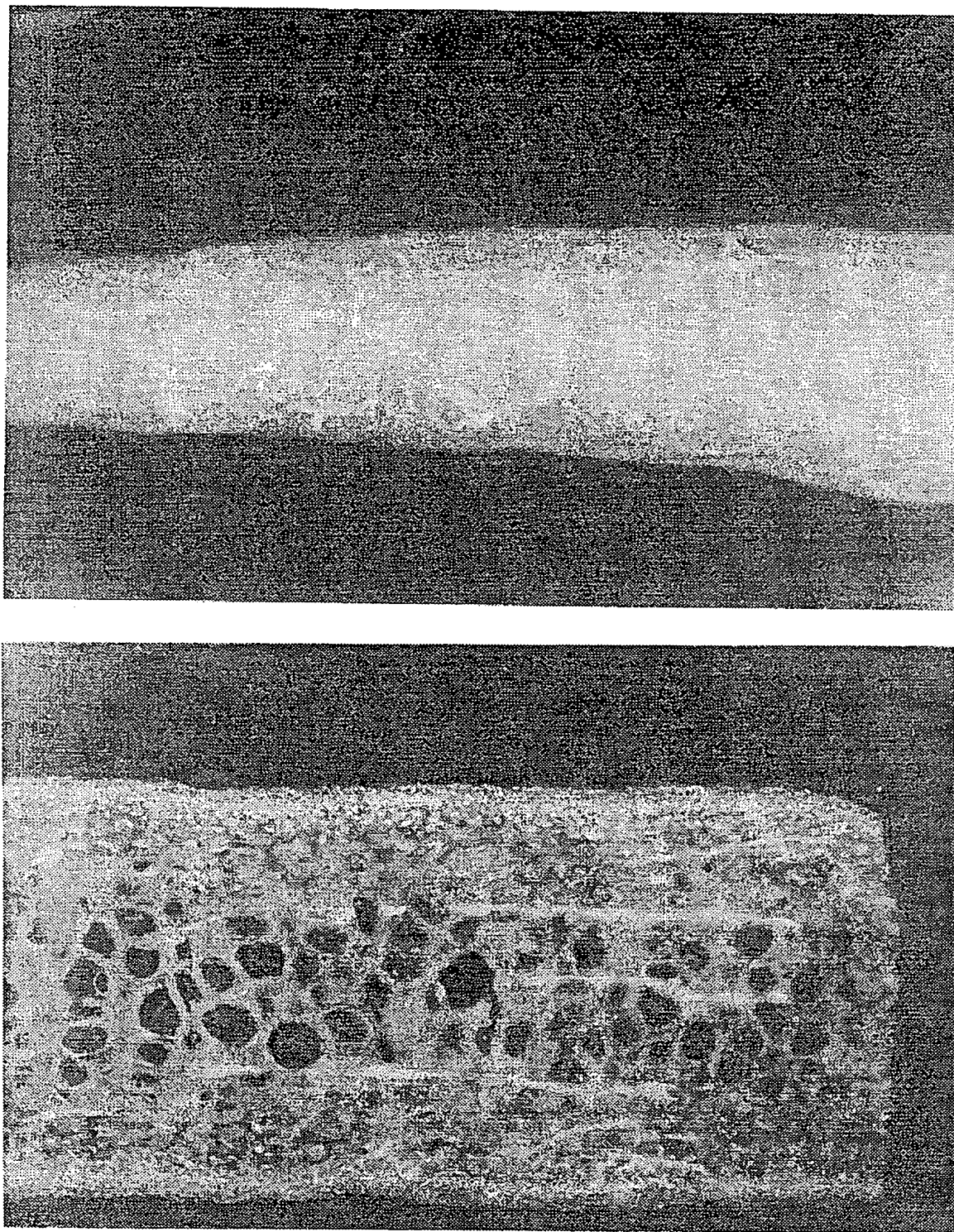


Fig.1. Calcofluor White stainings revealing fungi in upper part and lower part of *Gracilaria lemaneiformis*. (108x and 294x).

02 / 11

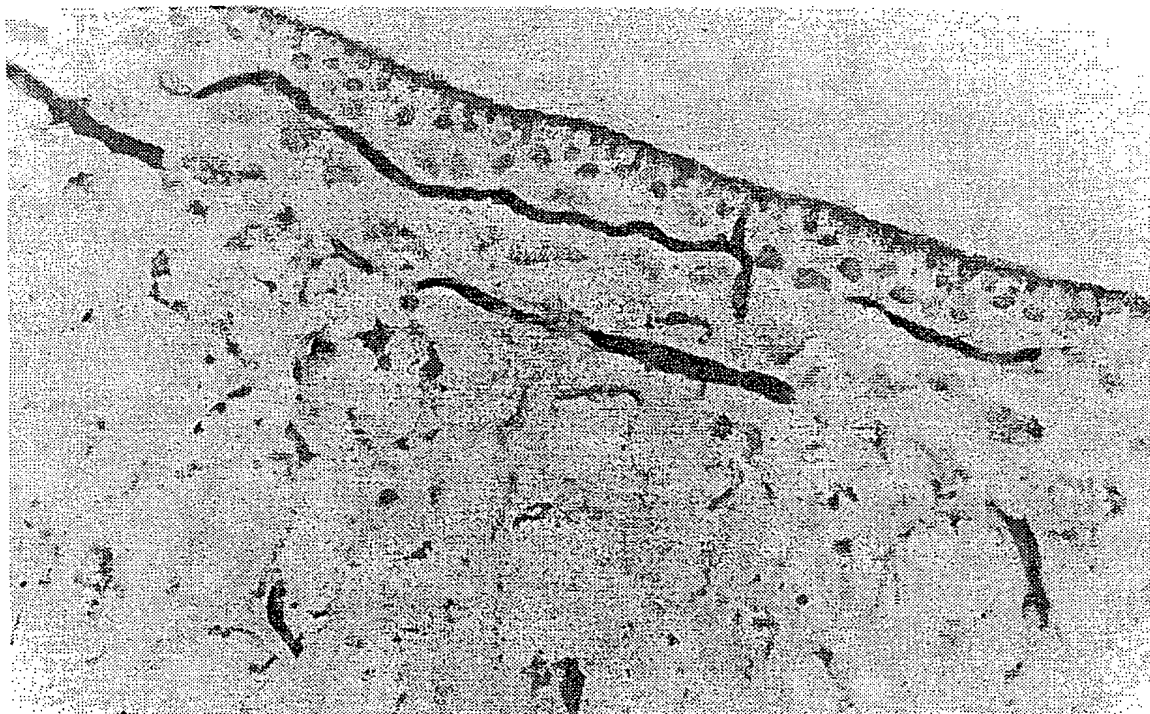


Fig.2. PAS / Anilinblue Black staining of *Gracilaria lemaneiformis* with fungi. The fungi have a significant higher content of carbohydrates.



Fig. 3. The micrograph shows longitudinal and grazing sections of two thin-walled fungal hypha (f) growing between thick walls (w) of algal cells. Note thylacoid membranes in the algal chloroplast (arrows).

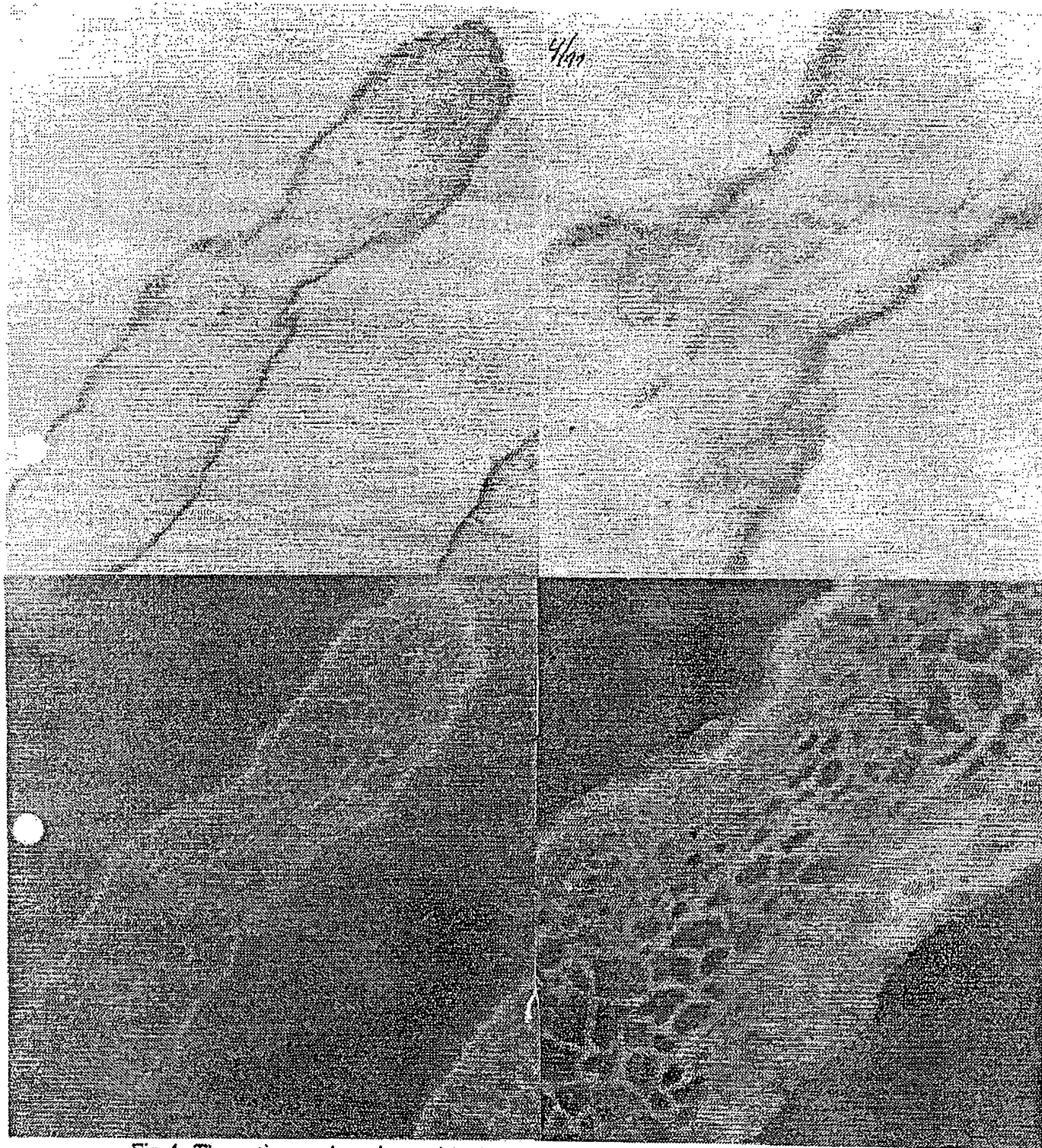


Fig.4. The antisense detections with clone 2 probe (upper row) are restricted to the fungi illustrated by the Calcofluor White staining of the succeeding section (lower row). (46x and 108x).

05 / 11

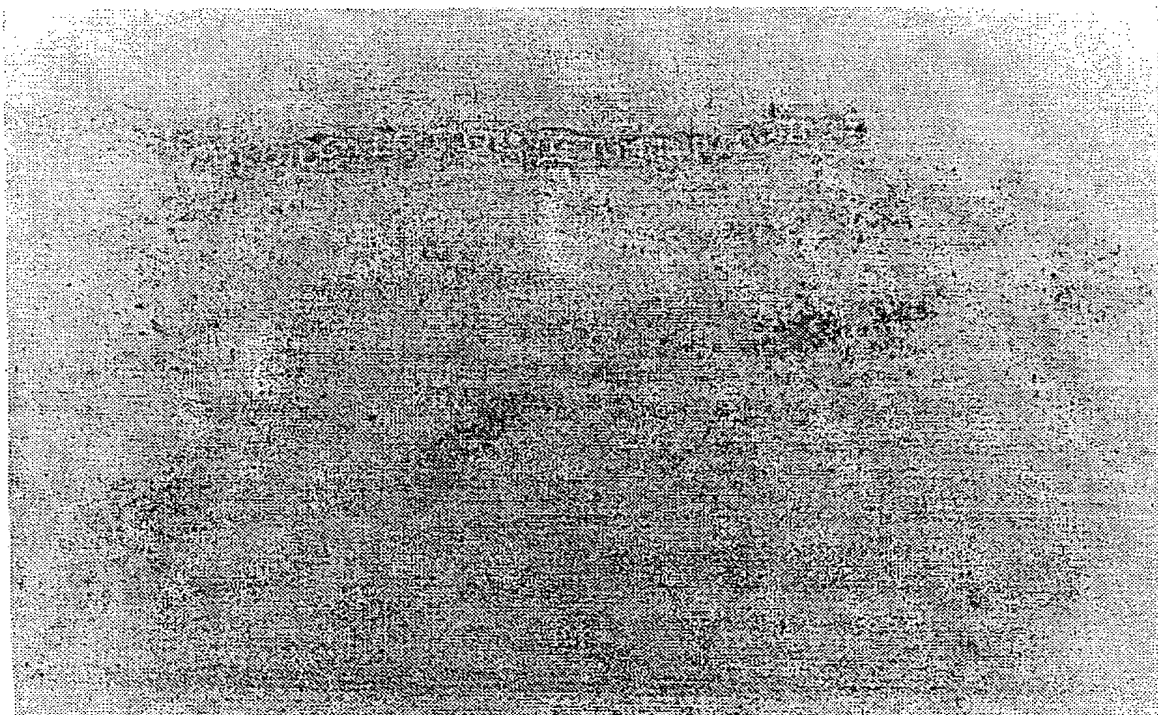
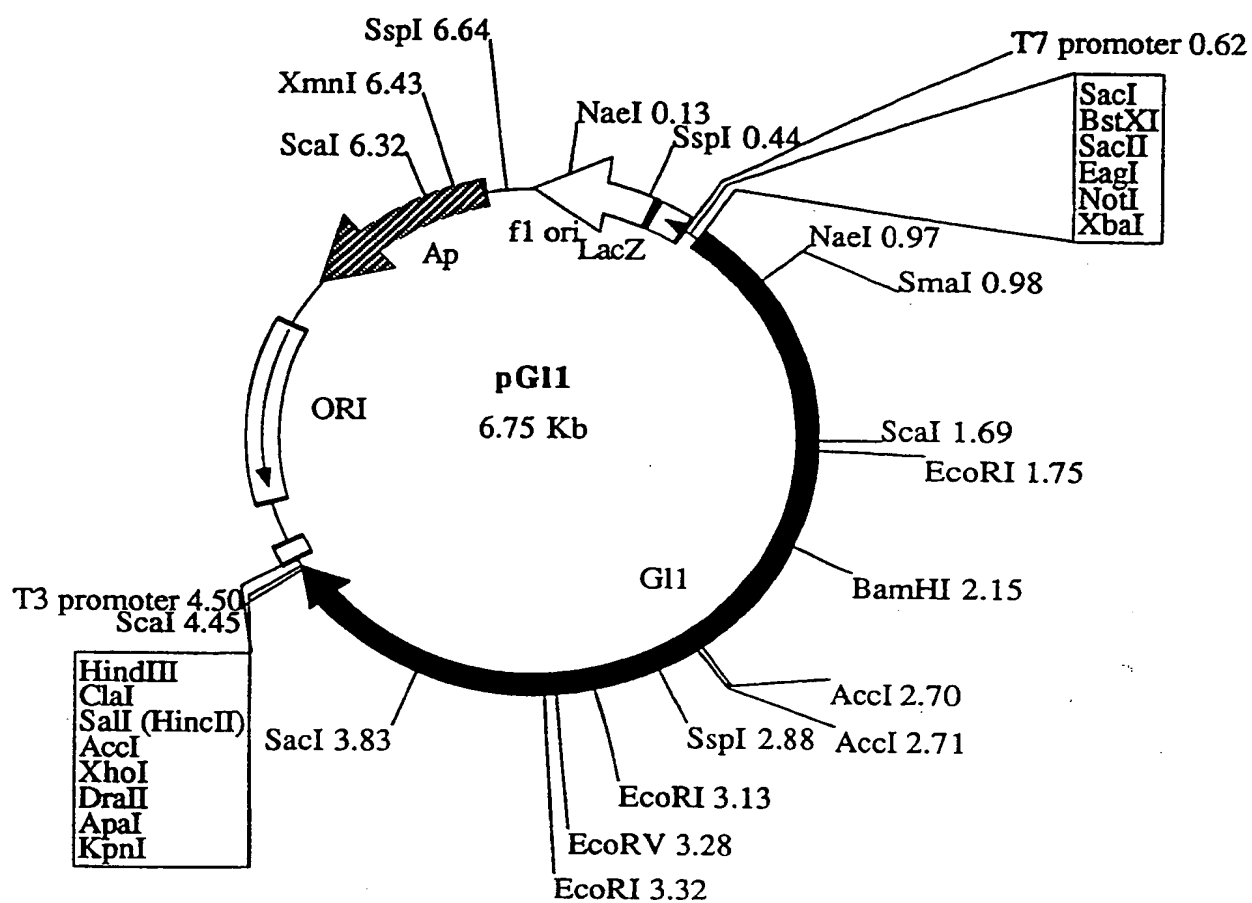


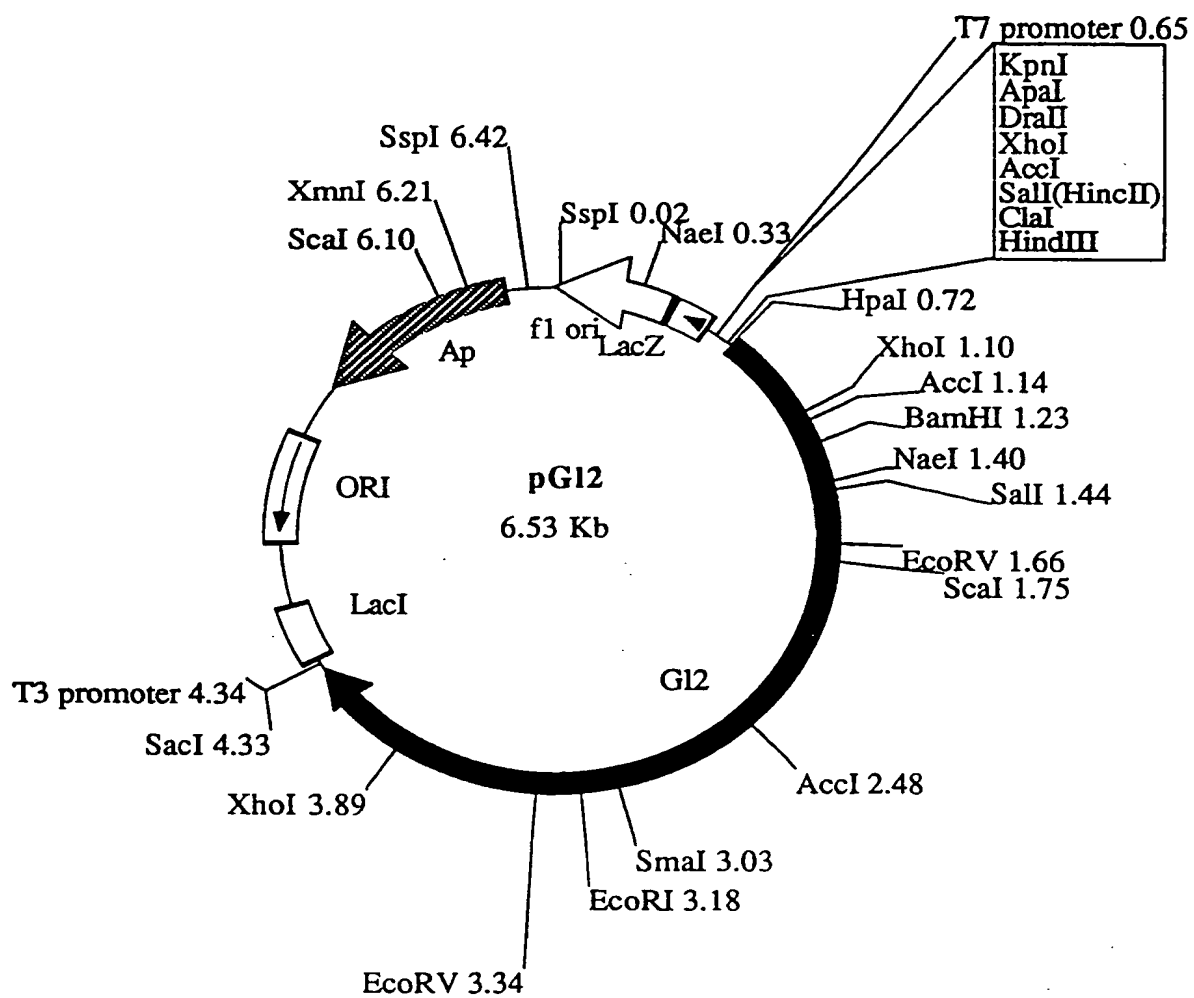
Fig.5. Intense antisense detections with clone 2 probe are found over the fungi in *Gracilaria lemnaeformis* (294x).

Fig. 6.



07 / 11

Fig. 7.



08 / 11

FIGURE 8

MFSTLAFVAP SALGASTFVG AEVRSNVRIH SAFAVHTAT RKTNRNLNVSM TALSDKQTAT
AGSTDNPDGI DYKTYDYVGV WGFSPLSNTN WFAAGSSTPG GITDWTATMN VNFDRIDNPS
ITVQHPVQVQ VTSYNNNSYR VRFNPDGPIR DVTRGPILKQ QLDWIRTQEL SEGCDPGMTF
TSEGFLTFET KDLSVIIYGN FKTRVTRKSD GKVIMENDEV GTASSGNKCR GLMFVDRLYG
NAIASVNKNF RNDAVKQEGF YGAGEVNCKY QDTYILERTG IAMTNYNYDN LNYNQWDLRP
PHHDGALNPD YYIPMYAAP WLIVNGCAGT SEQYSYGWFM DNVSQSYMNT GDTTWN SGQE
DLAYMGAQYG PFDQHFVYGA GGGMECVVTA FSLLOGKEFE NQVLNKR SVM PPKYVFGFFQ
GVFGTSSLLR AHMPAGENNI SVEEIVEGYQ NNNFPFEGLA VD VDMQDNLR VFTTKGEFWT
ANRVGTGGDP NNRSVFEWAH DKGLVCOTNI TCFLRNDNEG ODYEVNQT LR ERQLYTKNDS
LTGTD FGMTD DGPSDAYIGH LDYGGGVECD ALFPDWGRPD VAEWWGN NYK KLFSIGLDFV
WQDMTVPAMM PHKIGDDINV KPDGNWPNAD DPSNGQYNWK TYHPOVLVTD MRYENHGREP
MVTORNIHAY TLCESTRKEG IVENADTLTK FRRSYIISRG GYIGNQHFGG MWVGDNSTTS
NYIQMMIANN INMNMSCLPL VGSDIGGFTS YDNENQRTPC TGDLMVRYVQ AGCLLPWFRN
HYDRWIESKD HGKDYQELYM YPNEMD TLRK FVEFRYRWOE VLYTAMYONA AFGKPIIKAA
SMYNND SNVR RAONDHFLLG GHDGYRILCA PVVWENSTER ELYLPVLTOW YKFGPD FDTK
PLEGAMNGGD RIYNYPVPQS ESPIFVREGA ILPTRYTLNG ENKSLNTYTD EDPLVFEVFP
LGNNRADGMC YLDDGGVTTN AEDNGKFSVV KVAAEQDGGT ETITFTND CY EYVFGGPFYV
RVRGAQSPSN IHVSSGAGSQ DMKVSSATSR AALFNDGENG DFWVDOETDS LWLKLPNVVL
PDAVITIT


```
GL1      - YNNDNSVRRRAQNDHFLLLGGHDGYRILCAPVVWENSTERELYLPVLTLQWKYK -892  
          : :: ::: : : :::::::::::::::::::::::::::: : ::::: :  
GL2      - YDNDNRNVGAQDDHFLLLGGHDGYRILCAPVVWENTTSRDLYLPVLTKWKYK -895  
  
GL1      - FGPDPDTKPLEGAMNGGDRIYNYPVPQSESPIFVREGAILPTRYTLNGEN -942  
          : ::::: : . . : : : : : : : : : : : : : : : : : : : :  
GL2      - FGPDYDTRKRLDSALDGGQMKNYSVPQSDSPIFVREGAILPTRYTLDGSN -945  
  
GL1      - KSLNTYTDEDPLVFVEVPLGNNRADGMCYLDDGGVTNAEDNGKFSVVKV -992  
          : ::::: : : : : : : : : : : : : : : : : : : : : : : :  
GL2      - KSMNTYTDKDPLVFVEVPLGNNRADGMCYLDDGGITDAEDHGKFSVINV -995  
  
GL1      - AAEQDGGTETITFTNDCEYEVFGGPFYVRVGAQSPSNIHVSSGAGSQDM -1042  
          : : : : : : : : : : : : : : : : : : : : : : : : : :  
GL2      - EALRKGVTTTIKFAYDTYQYVFDGPFYVRIRNLTTASKINVSSGAGEEDM -1045  
  
GL1      - KVSSATSRAALFNDGENGDFFWDQETDSLWLKLPNVLPDAVITIT -1088  
          . : : : : : : : : : : : : : : : : : : : : : : : : :  
GL2      - TPTSANSRAALFSDGGVGGEYWADNDTSSLWMKLPNLVLQDAVITIT -1091
```

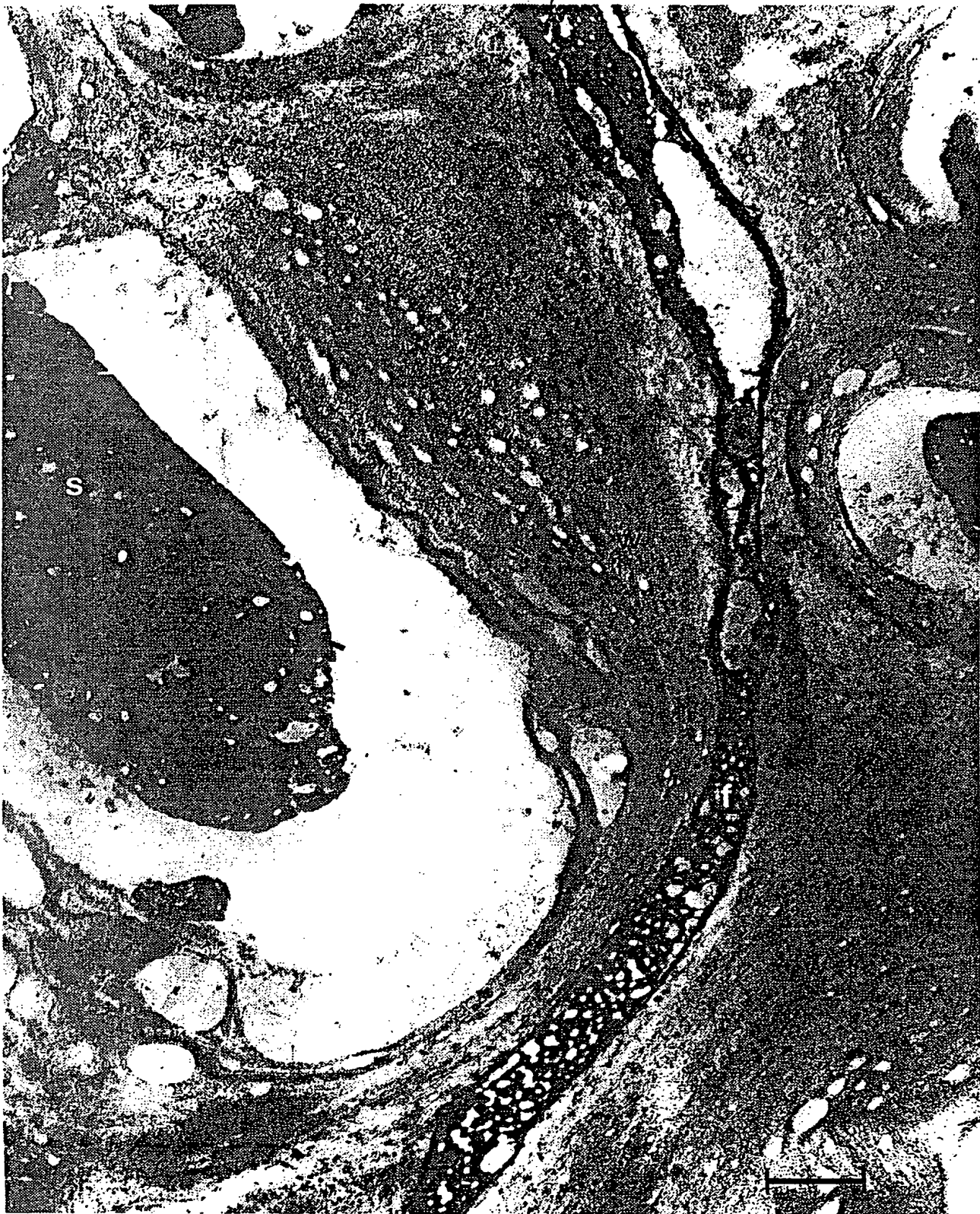


Figure 10. Microphotograph of a fungal hypha (f) growing between algal cell walls (w). Note grains of floridean starch (s) and thylakoids (arrows) in the algal cell. Bar = 2 μ m.